Development, Validation and Globalisation of a Health Status Measure for Evaluating Patients with Osteoarthritis

Nicholas Bellamy MB ChB, MD, MSc, MBA, FRCP (Glasgow), FRCP (Edinburgh), FRCP(C), FACP, FRACP, FAFRM
Professor of Rehabilitation Medicine and Director of the Centre of National Research on Disability and Rehabilitation Medicine, The University of Queensland, Brisbane, Australia

A thesis submitted to the University of Glasgow in fulfilment of the requirements for the Degree of Doctor of Science in Medicine

July 2005

Acknowledgements

I gratefully acknowledge the generous collaboration of the many research associates, technical experts, rheumatologists, orthopaedic surgeons, epidemiologists and biostatisticians who facilitated the successful development of the Western Ontario and McMaster (WOMAC) Osteoarthritis Index. These include senior academics in rheumatology who inspired the development and validation of the WOMAC Index, and colleagues in the private sector without whom, globalisation of the WOMAC Index would not have been feasible. I would particularly like to recognize Professor Watson Buchanan, Professor Peter Tugwell, and Professor Charles Goldsmith of McMaster University in Hamilton, Ontario, Canada, who in 1981 raised my interest in, and awareness of, deficiencies in outcome measurement in osteoarthritis (OA) clinical trials. Professor Watson Buchanan, in particular, provided considerable support for my early academic development, and encouragement in developing the WOMAC Index, and to him I am especially grateful.

The opportunity provided to me by the late Adam Linton, Professor of Medicine at The University of Western Ontario, and Chief of Medicine at Victoria Hospital, London Ontario, which allowed me to undertake an MSc degree at McMaster University, was the catalyst for a career in clinical research and my early academic development, and to him, I am also grateful. I also owe a debt of gratitude to Professor Hugh Simpson (University of Glasgow) and Professor Franz Halberg (University of Minnesota) for providing opportunity to participate in the data acquisition stage of a medical research project in NorthWest Greenland, and the analysis and write-up stages of that project at the University of Minnesota, while I was an undergraduate medical student at Glasgow University; an experience which highly influenced my subsequent choice of a career in medical research.

I am very grateful to my research assistant, Ms Jane Campbell, who has worked diligently over several years on many WOMAC research projects, and to my personal assistant, Ms Chesne McGrath, who has provided secretarial support since 1999 for the production of WOMAC publications, WOMAC questionnaires, and several editions of the WOMAC User Guide. I am extremely grateful to Dr Jennifer Kennedy who has assisted me, since 1999, in providing support to WOMAC users, meeting their measurement needs, and answering their questions and correspondences.

I am also grateful to Professor Brooks, Dean of the Faculty of Health Sciences at The University of Queensland, Brisbane, Australia, for the opportunity, since 1999, to continue my research into health status assessment in OA as integral to my appointment as Professor of Rehabilitation Medicine and Director of the Centre of National Research on Disability and Rehabilitation Medicine at The University of Queensland.

Finally, I am especially grateful to the many patients with OA, who I have had opportunity to interview, and whose careful explanation of the impact of OA on different aspects of their everyday lives, shaped the content and format of the WOMAC Index. Without their insight and perseverance, it would not have been possible to successfully develop a valid health status measure meeting a global need.

LIST OF CONTENTS

SUMMARY

The data acquisition, analysis and interpretation, which inform the text of this thesis, were conducted in the period after the award of my MSc (McMaster)(1982) and this DSc thesis submission (2005). In 1982 there existed only rudimentary guidelines for the conduct of osteoarthritis (OA) clinical trials, few health status measures suitable for performing measurement in rheumatology environments, no agreement on core set measures, or preferred instruments, and no responder or state attainment criteria. My MSc thesis was a design thesis, in which I had conducted a review of the OA clinical trials and outcome measurement literatures and proposed a theoretical construct for the development of an evaluative index for OA clinical trials. That review confirmed a high degree of heterogeneity in outcome measurement methods, a paucity of standardised methods for assessing disease-specific health-related quality of life, and a lack of international consensus on core set measurement requirements and preferred methods. Following my MSc graduation I developed a programme of research to develop, validate and globalise a valid, reliable and responsive standard of measurement (Western Ontario and McMaster Osteoarthritis Index – WOMAC Index) for OA clinical trials. This programme was initiated at The University of Western Ontario (UWO) (1982-1999) and continued at The University of Queensland (UQ) (1999-2005). In addition to my own research programme, which I led as Professor of Medicine at UWO and subsequently as Professor of Rehabilitation Medicine at UQ, I also collaborated with other research groups, organisations and agencies to progress international research agendas in OA clinical research methodology, many of which were based on WOMAC data.

The initial phase (1982-1992) of WOMAC development involved development and validation. Specification of the item content was achieved through face-to-face interview of 100 patients with hip and/or knee OA. The resulting test index was composed of five subscales. Two independent validation studies, involving two different scaling formats, were designed and executed, one in an orthopaedic environment involving total joint arthroplasty, and the other in a rheumatology environment involving a double-blind randomised controlled clinical trial of two nonsteroidal anti-inflammatory drugs (NSAIDs). Four of the five subscales were successfully validated, of which three were retained in the final Index. The face, content and construct validity, reliability and responsiveness of the WOMAC Index were established. From these and other early studies several other performance-based features of the Index were characterized. These included the effects of prior score availability versus non-availability, time frame variations from 24 hours to two weeks, relative responsiveness versus other measures, selecting signal items versus using the entire Index, and parametric versus non-parametric analysis.

The intermediate phase (1993-1999) of WOMAC development involved globalisation, and occurred *pari passu* with international attempts to harmonise OA clinical trials methodology; evidence-based and consensus processes that I was involved in through OMERACT and OARSI Task Force participation. Most importantly for the WOMAC Index were growing opportunities to develop and validate alternate-language translations of the Index, based on the original WOMAC 3.0 Index English for Canada source questionnaire. The resulting 32 linguistically validated alternate-language translations, eventually increased to over 60 alternate-language forms, most being

available in both adjectival and visual analogue scaling formats. The end result of the aforementioned processes was the globalisation of the WOMAC Index, international consensus on core set domains for OA outcome measurement and specification of preferred measures, one of which was the WOMAC Index. Rapidly expanding utilisation of the WOMAC Index by academically-based and industry-based researchers, was shortly thereafter followed by a sharp increase in the number of studies reporting use of the WOMAC Index, such that by 1999 it was often the most commonly used health status questionnaire in osteoarthritis clinical research reported at major rheumatology conferences in Europe, N. America and Australasia.

The late phase of development (2000-2005) has involved the further development of other language forms, other scaling formats, short forms and versions amenable to telephone administration and electronic data capture. This phase has also involved using WOMAC Index data to facilitate the development, by various research groups with whom I have collaborated, of definitions of responder criteria and state-attainment criteria. In particular, we have used WOMAC data, in whole or part, in the development of the following definitions of responder criteria: OARSI responder criteria, OMERACT-OARSI responder criteria, Minimum Perceptible Clinical Improvement (MPCI), Minimal Clinically Important Improvement (MCII), and in the development of the following definition of state-attainment criteria: Patient Acceptable Symptom State (PASS).

The now fully developed WOMAC Index is a tri-dimensional, disease-specific, self-administered, health status measure. It probes clinically-important, patient-relevant symptoms in the areas of pain, stiffness and physical function in patients with OA of the hip and/or knee. The index consists of 24 questions (5 pain, 2 stiffness, 17 physical function) and can be completed in less than 5 minutes. It is available in Likert (WOMAC LK-series), Visual Analogue (WOMAC VA-series) and Numerical Rating (WOMAC NRS-series) scaled formats. WOMAC is valid, reliable, and sufficiently sensitive to detect clinically-important changes in health status following a variety of interventions (pharmacologic, surgical, physiotherapy, etc). It has been translated into many different languages and has been requested for use by more than 500 researchers in over 50 different countries. Several different formats of the WOMAC Index have been produced including the WOMAC 3.0, 3.0S, 3.1, 3.1S, 3.1W, 3.1M and 3.1SLV, 3.1IK, SF-WOMAC and e-WOMAC.

The WOMAC Index has become a global standard of measurement for clinical trials in hip and knee OA in rheumatology, is widely used in clinical research, and has been incorporated into several major regulatory and guidelines documents. The WOMAC Index has been important to the development of global harmonisation in outcome measurement, in formulating response and state attainment criteria, and in adjudicating the clinical benefit of new treatments for knee OA.

SECTION 1 – Introduction

Following the successful completion, in June 1982, of an MSc in Clinical Epidemiology and Biostatistics at McMaster University in Hamilton, Ontario, Canada, and subsequent relocation to a faculty position at The University of Western Ontario in London, Ontario, Canada, I embarked on a series of studies conducted between late-1982 and mid-2004 that resulted in the development of an internationally recognised, and widely used, measurement tool called the Western Ontario and McMaster Osteoarthritis (WOMAC) Index.

By completion of my MSc degree in 1982, I had recognised a lack of standardisation in outcome measurement procedures in osteoarthritis (OA) research publications in rheumatology, as well as a paucity of detailed guidelines specific to the conduct of outcome measurement in OA clinical trials of pharmacologic agents. In particular, I had recognised that in clinical trials reports not only did the variables measured greatly differ, but there was considerable variability in the scales and instruments employed to capture data on even a single variable, for example pain. Furthermore, apart from physician and patient global assessments, pain and stiffness were the only other two variables monitored in more than half of the 63 trials critically appraised, while the second most important symptom of OA, physical disability, was monitored in only 35% of the studies. The European League Against Rheumatism (EULAR) OA clinical trails guidelines only existed in draft form in 1982, and the United States Food and Drug Administration (FDA) guidelines gave relatively little guidance regarding outcome measurement in OA clinical trials, the latter recommending measurement of joint range of motion and walking or stair climbing time, but not specifically recommending measurement of physical functioning or disability.

The need for a standardised method of capturing patient-centred outcomes or patient reported outcomes (PROs) was timely and attainable. Following completion of my MSc degree, I relocated from Hamilton, Ontario to London, Ontario in July 1982, to take up faculty appointments, at the Assistant Professor level, in Medicine, Epidemiology and Biostatistics at The University of Western Ontario and a consultancy in rheumatology at Victoria Hospital. I also retained a part-time faculty level appointment at McMaster University as Assistant Professor of Clinical Epidemiology and Biostatistics.

Soon after arriving in London, I initiated the first studies directed at developing the WOMAC Index. This thesis describes the research work undertaken in Ontario, Canada, and after 1999, in Queensland, Australia, that resulted in the development, validation and globalisation of the WOMAC Index, a patient-centred health status questionnaire, now cited in international guidelines for outcome measurement in clinical trials, widely used globally in 65 languages in clinical research environments in rheumatology, and which has played an important role in setting standards for proposed response and state-attainment criteria.

Following this Introduction, the thesis is divided into nine other sections as follows: Development, Validation, Exploration of Special Measurement Characteristics, Applications in Clinical Research, Globalisation, Flexible Delivery, Responder Criteria and State-Attainment Criteria, WOMAC Index: A Global Perspective, and WOMAC Index: Contemporary Context. The sections follow a basic sequence from the earliest stage of development through to the most recent publications, and within each section

and subsection are generally divided according to chronology. All except the WOMAC User Guide have been previously published in peer review journals, and represent important contributions to the history of outcome measurement development in OA, and the WOMAC Index development in particular.

Section 2 – Development, describes the patient-based method used to develop the item inventory for the WOMAC Index, a process that conferred the necessary face and content validity to the Index. Section 3 – Validation, contains a review of the two major validation studies that confirmed the construct validity, test-retest validity, internal consistency and responsiveness of the WOMAC Index, and resulted in the selection of three of the original five dimensions for further development. Section 4 - Exploration of Special Measurement Characteristics, reviews sub-analyses and sub-studies conducted to explore issues of blind versus informed presentation, signal vs aggregate strategies of measurement, time frame dependency of responses to the WOMAC Index and the relative responsiveness of the WOMAC Index compared to other methods of outcome assessment. Section 5 - Applications in Clinical Research, describes post-validation experience with the WOMAC Index in different research environments including NSAIDs, a complex analgesic, a slow-acting drug for osteoarthritis and a viscosupplement. Section 6 – Globalisation, describes contributions made, in collaboration with working groups and task forces of the Outcome Measures in Rheumatology Clinical Trials (OMERACT) Group, Osteoarthritis Research Society International (OARSI), World Health Organisation / International League of Associations of Rheumatology (WHO/ILAR) and the Initiative on Methods, Measurement and Pain Assessment in Clinical Trials (IMMPACT) Group, to the development of international consensus and harmonisation in outcome measurement in OA clinical trials, developments in which the WOMAC Index played an important role. The section also details the methods used to successfully globalise the WOMAC Index in over 60 alternate-language translations. Section 7 - Flexible Delivery, describes collaborative projects directed at validating the WOMAC Index for delivery by telephone interview, and for electronic data capture using mouse-driven cursor and touch-screen technology on personal computers, as well as documenting the validation of a short form of the WOMAC Index. This section also describes, based on surveys undertaken in Canada and Australia, the limited use in clinical practice of quantitative methods of health status assessment, which were frequently used in clinical research environments. Section 8 - Responder Criteria and State-Attainment Criteria, describes the contributions made by the author and the role of the WOMAC Index in the development of proposed definitions for responder and state-attainment criteria in OA. Section 9 - WOMAC Index: A Global Perspective, reviews the current status of the WOMAC Index, and summarises contributions made by the author and the Index, in the development of global opportunities for standardised health status measurement in OA in rheumatology. Section 10 - WOMAC Index: Contemporary Context, describes the placement of WOMAC Index, within the context of other measures for hip and knee OA in rheumatology, that have emerged subsequent to validation of the WOMAC Index.

A list of the publications that contribute to this thesis, is provided after the ten aforementioned sections. Finally, copies of the WOMAC Copyright Certificate and the WOMAC Trademark Certificate issued to me, by the Government of Canada, are provided in Appendix A, and OA measurement alternatives are listed in Appendix B.

SECTION 2 - Development

My research goal was to develop a valid, reliable and responsive patient self-reported health status questionnaire, to meet a global measurement need in clinical trials of pharmacologic agents in hip and knee OA. Therefore, in order to construct the item inventory of WOMAC, the dimensionality of the symptomatology of hip and knee OA was explored in 100 patients with hip and/or knee involvement (1). The survey questionnaire was developed by a peer review process involving the opinions of four academic rheumatologists and two clinical epidemiologists experienced in clinical measurement in the rheumatic diseases. Initial questions were open-ended and probed the clinical importance and characteristics of any pain, stiffness, physical, social or emotional dysfunction. Once spontaneous responses to these questions were exhausted, a battery of closed-ended questions, derived from (and modified where necessary) six existing questionnaires (Health Assessment Questionnaire – HAQ, Functional Status Index – FSI, Arthritis Impact Measurement Scales – AIMS, Pooled Index, McMaster/Toronto Assessment Index – MACTAR, McMaster Health Index Questionnaire – MHIQ), was used to complete the assessment of each dimension and quantify any sources of discomfort and disability. The survey questionnaire was administered by face-to-face interview. The following data were recorded: 1) the presence or absence of each of several types of discomfort or disability, 2) the frequency with which each type of discomfort or disability occurred (daily, weekly, fortnightly, monthly or less), and 3) the importance of each type of discomfort or disability to the patient (0 = none, 1 = slight, 2 = moderate, 3 = very, 4 = extreme).

Patients were specifically questioned about sources of discomfort and disability recently experienced, and attributed to OA in the hips and/or knees. They were questioned regarding the perceived importance of each type of discomfort and disability, in order to assess their clinical relevance. Gender-specific questions relating to physical disability (e.g. ironing) were avoided, and questions phrased in more general terms (e.g. light domestic duties). Questions relating to sexual function were not included, since this had been previously noted to result in low response rates, and inhibit responses to subsequent non-sexual function questions.

Using this approach the item inventory for a prototype WOMAC Index was specified, based on the prevalence, frequency and mean importance of reported symptoms on each of the five dimensions (pain, stiffness, physical function, emotional function, social function). A prototype WOMAC Index was prepared in both 5-point adjectival (syn:Likert or LK) and 100 mm visual analogue (VA) scaling formats. The questions included in the prototype WOMAC Index inventory targeted symptoms, that patients generally considered important, and which recurred with relative frequency.

This method of development, based on literature review, consultation with key informants and structured face-to-face interviews with 100 patients with hip and/or knee OA, using both closed and open-ended questions, was considered sufficient to confer the necessary face and content validity to the prototype WOMAC Index. Indeed this patient-based approach is considered a key determinant of the global success of the WOMAC Index, and differentiates it from health status measurement tools based only on expert opinion. While expert opinion can be informative, it may not adequately reflect the patient's perspective, and may be at risk of being considered paternalistic. In contrast, the

approach used in the development of the WOMAC Index was predominantly driven by the opinion and experience of patients with symptomatic hip and/or knee OA.

Reference:

1.Bellamy, N. and Buchanan, W.W. A preliminary evaluation of the dimensionality and clinical importance of pain and disability in osteoarthritis of the hip and knee. Clinical Rheumatology 1986;5(2):231-241.

Clinical rheumatology, 1986, 5, N° 2 231-241

A preliminary evaluation of the dimensionality and clinical importance of pain and disability in osteoarthritis of the hip and knee

N. BELLAMY, W.W. BUCHANAN

Department of Rheumatology, University of Western Ontario, London, Canada; Department of Rheumatology, McMaster University, Hamilton, Ontario, Canada.

SUMMARY *Current methods of clinical assessment in osteoarthritis show a high degree of variability. By contrast, patients with rheumatoid arthritis may be evaluated using a number of standardised and validated indices. One hundred patients with primary osteoarthritis of the hip and knee were interviewed in order to determine the dimensionality of their discomfort and disability and to define the clinical importance of each component item. The symptomatology of osteoarthritis was captured by five pain, one stiffness, twenty-two physical, eight social and eleven emotional items. In spite of a high degree of variability in the frequency of involvement of the individual items, their clinical importance was similar both within as well as across dimensions. Further studies are indicated to determine the reliability, validity and responsiveness of each of the items identified as a prelude to developing a standardized method of assessing patients with osteoarthritis of the hip and knee.*

Key words: Osteoarthritis, Pain, Disability, Clinical Importance.

INTRODUCTION

Discomfort (pain and stiffness) and disability (physical, social and emotional) are the major symptoms of osteoarthritis. In spite of the greater prevalence of osteoarthritis, more attention has been directed towards the study of functional decrements and the quality of life of patients with rheumatoid arthritis (1) than to patients with degenerative forms of arthritis (2,3). In respect to outcome measurement in osteoarthritis clinical trials, we have reported in a previous edition of this journal, a review of 63 clinical studies of nonsteroidal anti-inflammatory drugs reported between 1962 and 1982 and have observed a high degree of variability in the outcome measures employed (4). In addition to lacking any standardisation, current measures presume a validity extrapolated from the rheumatoid arthritis literature. Thus, the majority of indices which have been developed for use in rheumatic diseases have been based on patients with rheumatoid arthritis (5-29). However, fundamental differences exist between patients with rheumatoid and osteoarthritis in respect of the age of onset, distribution of joint involvement, natural history of the disease and response to treatment. Only the Doyle (7) and Lequesne (24) indices have

Received 18 March 1985,
Revision - accepted 10 February 1986
Correspondence to: DR. N. BELLAMY
Suite 402 A, Victoria Hospital (Westminster Tower), 800 Commissioners Road East London, Ontario, Canada, N6A 4G5

been expressly developed for evaluating patients with osteoarthritis. However, the Doyle Index is unidimensional and is a modification of the Ritchie Index while the Lequesne Index is oligodimensional and utilizes a restricted number of response alternatives.

In view of these deficiencies in outcome measurement in osteoarthritis clinical trials, we are currently undertaking a series of studies to rationalize the measurement process pertaining to patients with primary osteoarthritis of the hip and/or knee. In the present study the extent of each of five content domains was assessed and component items ranked according to their prevalence and clinical importance. The objective was to define the dimensionality of pain and disability and identify those component items having the greatest clinical importance in a group of potential drug-study patients.

MATERIALS AND METHODS

One hundred out-patients with osteoarthritis of the hip and/or knee were selected for study. To be eligible patients had to fulfill the following criteria: 1) Attend a rheumatological clinic at either the University of Western Ontario, London, or McMaster University Medical Centre, Hamilton; 2) Be ambulatory; 3) Have symptomatic primary osteoarthritis affecting at least one hip or knee and requiring treatment with a nonsteroidal anti-inflammatory analgesic medication; 4) Have minimal or no spinal symptoms; 5) Be unrestricted (in their functional capacity) by any co-morbid condition and; 6) Not have had prior hip or knee replacement surgery or an osteotomy. The patients selected for study would all have been eligible for a clinical trial of nonsteroidal anti-inflammatory drug therapy since they were all typical of patients commonly used in such trials.

The survey questionnaire was developed by a peer review process utilizing the opinions of four rheumatologists (WWB, NB, PT, PJR) and two clinical epidemiologists experienced in clinical measurement in the rheumatic diseases (CG, LC). Initial questions were open-ended and probed the clinical importance and characteristics of any pain, stiffness, physical, social or emotional dysfunction. Once spontaneous responses to these questions were exhausted, a battery of closed-ended questions derived from six existing questionnaires (10-13, 25, 27, 31, 32) was used to complete the assessment of each dimension and quantitative any sources of discomfort or disability detected.

The following data were recorded: 1) The presence or absence of each of several types of discomfort or disability (Table I-IV); 2) The frequency with which each type of discomfort or disability occurred (daily, weekly, fortnightly, monthly or less) and 3) The importance of the discomfort or disability to the patient (0 = none, 1 = slight, 2 = moderate, 3 = very, 4 = extremely). It should be noted that patients were specifically asked to record the perceived importance of each type of discomfort or disability reported in order to assess its clinical relevance. Furthermore, the discomfort and disability sought was specified as having been recently experienced and directly related to osteoarthritis of the hip and/or knee. Thus, each patient was asked to report only those symptoms which they felt were the direct result of their articular disease.

During questionnaire construction, items directed specifically at patients of one or other sex (e.g. ironing) were avoided and the questions rephrased in more general terms (e.g. light domestic duties). Patients were not asked about sexual function in order to avoid embarrassment and because this has been previously noted to inhibit responses even to subsequent non-sexual questions.

Before being formally applied the questionnaire was pre-tested in 15 osteoarthritic patients in order to assess its comprehensibility and feasibility. Thereafter, the questionnaire was administered to 90

patients by face-to-face interview (using trained interviewers) and to a further 10 patients by telephone. Telephone interviews were permitted in order to be able to survey patients, at either end of the severity spectrum, who though ambulatory did not wish to make a non-essential journey. Patients who were confined to a bed or wheelchair were excluded from the survey as they would not normally have been admitted to a drug trial. The response rate amongst those invited to participate was 97%.

Table I *Pain rank ordered by prevalence (P)*

ITEM	P	MIS	[DW%]
Walking	.77	2.58	96
Stairs	.75	2.62	94
In bed	.67	2.63	96
Weight bearing	.57	2.51	94
Sitting/lying	.56	2.57	95
Bending	.01	3.00	100
Strenuous exercise	.01	3.00	100

Table II *Physical disability rank ordered by prevalence (P)*

ITEM	P	MIS	[DW%]
Rising from sitting	.70	2.32	99
Descending stairs	.69	2.60	94
Ascending stairs	.69	2.54	89
Standing	.57	2.64	96
Walking on flat surfaces	.56	2.40	96
Getting in/out of car	.56	2.26	91
Bending to floor	.55	2.51	95
Going shopping	.49	2.40	94
Going shopping	.49	2.40	94
Putting on socks	.46	2.38	96
Rising from bed	.45	2.37	100
Taking off socks	.43	2.37	95
Getting in/out of bath	.40	2.30	98
Lying in bed	.39	2.36	95
Heavy domestic duties	.36	2.43	71
Light domestic duties	.36	2.26	88
Sitting	.35	2.54	100
Getting on/off toilet	.33	2.67	85
Getting on/off bus	.30	2.38	66
Getting in/out shower	.16	2.31	94
Driving a car	.16	2.25	75
Going from bed to chair	.15	2.21	85
Running on flat surfaces	.13	2.42	75

Table III *Social function rank ordered by prevalence (P)*

ITEM	P	MIS	[DW%]
Restricted leisure activities	.54	2.56	87
Attendance at community events	.27	2.15	77
Attendance at church	.23	2.52	74
Relations with spouse	.20	2.65	90
Relations with family	.18	2.67	89
Relations with friends	.14	2.64	79
Relations with others	.11	2.55	91
Dancing	.03	2.33	33

Table IV *Emotional function rank ordered by prevalence (P)*

ITEM	P	MIS	[DW%]
Frustration	.56	2.44	86
Anxiety	.55	2.64	74
Irritability	.53	2.59	81
Depression	.45	2.49	65
Difficulty relaxing	.44	2.39	73
Difficulty sleeping	.36	2.58	98
Boredom	.26	2.62	88
Loneliness	.23	2.26	78
Difficulty coping with stress	.17	2.19	75
Disturbed sense of well-being	.14	2.62	54
Poor self-control	.09	2.11	78

Following completion of 100 interviews, the data were summarized to provide the following values: 1) Prevalence of each type of discomfort or disability (P); 2) Mean importance score (MIS = the sum of the individual importance scores given by n affected individuals divided by n; and 3) The percentage of symptomatic patients experiencing daily or weekly symptoms (DW% = High Frequency). The individual items were then ranked within each dimension in order of their prevalence. In this study prevalence was defined as the proportion of patients in the "at risk" population who were concerned by ongoing symptomatology on a given variable (Tables I-IV). Subsequent analyses examined the effects of age, sex and disease

duration on the key symptoms within each dimension (Tables V and VI, Fig. 1-3).

RESULTS

Sixty-three female and 37 male patients were interviewed. The mean age was 61.07 years (range = 27-93) and mean disease duration (i.e. symptomatic) 10.07 years (range = 0.25 - 51). Eleven patients had hip disease alone, 57 knee disease alone, and in 32 the disease affected both hip and knee. Eight patients had previously undergone menisectomy some years earlier but none had been subject to arthroplastic surgery or osteotomy. All patients were symptomatic at the time of assessment.

Pain

Pain was disaggregated into pain occurring during five types of activity (Table I). Only two additional components of pain were identified by open-ended questions and it can therefore be assumed that these five principle items adequately represent the dimension of pain. Only one patient complained of pain while bending from the waist and a second of pain during strenuous exercise. The prevalence of different sources of pain varied from 56 to 77%, walking being the most frequent cause of pain. Amongst the five principle items, pain on sitting or lying was the least frequent (56%). Mean importance (MI) scores for pain varied from 2.509 to 2.627, and pain occurred with high frequency, being present at least daily or weekly in 94-96% of affected patients.

Pain prevalance for pain in bed and while negotiating stairs showed little variation with age but there was a tendency for pain experienced while sitting and while walking to increase prevalence with advancing age (Fig. 1). Overall MI scores (Fig. 1) failed to show any correlation ($r = 0.00$) with age (Table V) and there was no significant difference ($p = 0.88$) in MI scores for pain between males and females (Table VI). There was a low level of correlation ($r = 0.26$) between MI scores for pain and disease duration and there were modest but significant correlations between MI scores for pain and those for stiffness, physical

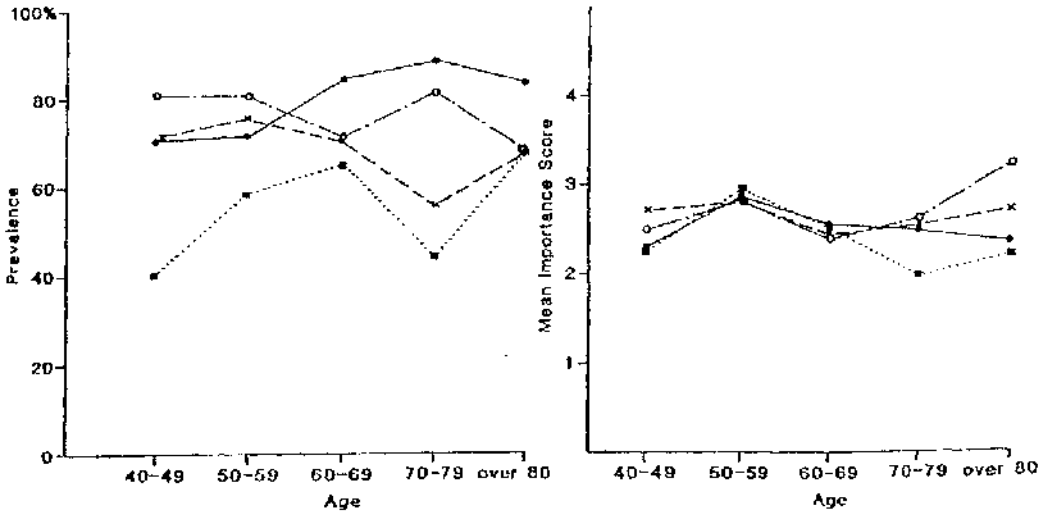


Fig. 1: Prevalence and mean importance scores for selected sources of pain as a function of age (• = Walking, o = Negotiating stairs, x = In bed at night, ■ = sitting or lying).

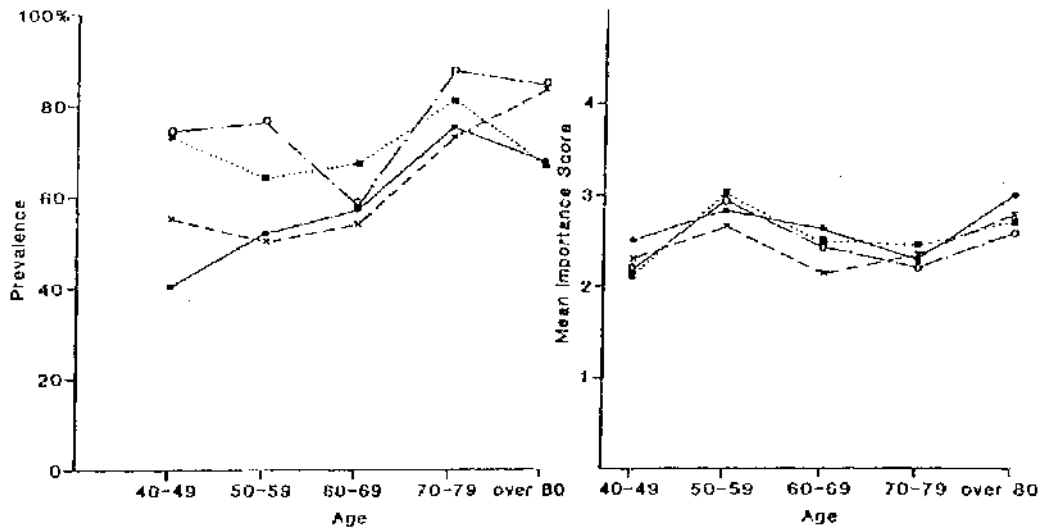


Fig. 2: Prevalence and mean importance scores for selected physical disabilities as a function of age (• = Standing, o = Ascending stairs, x = Walking on a flat surface, ■ = Descending stairs).

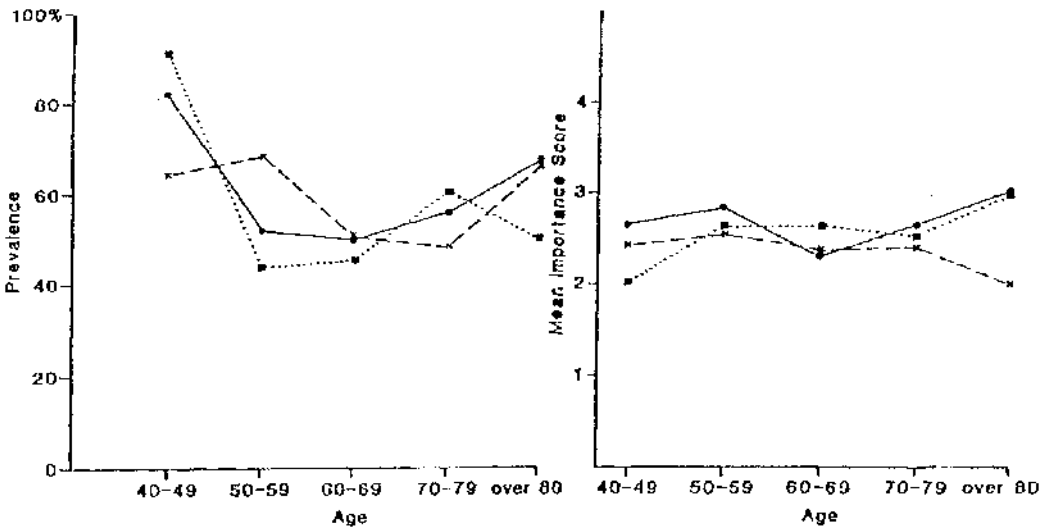


Fig. 3: Prevalence and mean importance scores for selected socioemotional disabilities as a function of age (• = Anxiety, x = Frustration, ■ = Restricted leisure activities).

dysfunction, and key social and emotional items (Table V).

Stiffness

Forty-seven percent of patients complained of joint stiffness after wakening in the morning and 73% after prolonged sitting or lying at other times. Mean importance scores were 2.524 and 2.303 respectively. The mean duration of morning stiffness was 14.5 minutes (range 1-120). Stiffness occurred on a daily basis in almost all affected individuals.

Table V *Correlation matrix for mean importance scores for pain, stiffness, physical, social and emotional function, age and disease duration*

	Pain	Stiffness	Physical function	Social function	Emotional function			Age	Disease Duration
					Anxiety	Frustration	Irritability		
Pain	1.00								
Stiffness	0.21	1.00							
Physical	0.74	0.22	1.00						
Social	0.39	0.23	0.44	1.00					
Anxiety	0.44	0.23	0.52	0.49	1.00				
Frustration	0.46	0.23	0.50	0.41	0.59	1.00			
Irritability	0.40	0.16	0.40	0.21	0.57	0.55	1.00		
Age	0.00	-0.03	0.06	-0.14	-0.14	-0.13	-0.20	1.00	
Disease duration	0.26	-0.01	0.08	-0.03	0.14	-0.01	-0.02	0.14	1.00

Table VI *Comparison of the mean importance scores given by symptomatic male versus female patients for each dimension*

	Pain	Stiffness	Physical	Social	Emotional
Male	2.56	2.28	2.16	2.38	2.44
Female	2.52	2.32	2.38	2.23	2.35
p value (2-tailed)	0.88(NS)	0.84(NS)	0.22(NS)	0.43(NS)	0.51(NS)

No significant correlation was noted between MI scores for stiffness and either age ($r = -0.03$) or disease duration ($r = -0.01$) (Table V). Low levels of correlation were detected between stiffness scores and those of pain, physical function, restricted leisure activity, anxiety and frustration (Table V). Although morning stiffness was more prolonged in females (mean = 13.55 minutes) than males (mean = 7.42 minutes) this difference was not statistically significant ($p = 0.24$). Similarly, no significant differences ($p = 0.84$) were detected in MI scores for stiffness between males and females (Table V).

Physical Dysfunction

Physical dysfunction was disaggregated into disability occurring during 22 types of activity (Table II). No additional components of physical disability were identified by open-ended questions and therefore these 22 questions are considered as adequately representing this dimension. The prevalence of physical disability varied from 13 to 70%. Negotiating stairs, rising from the seated position, standing, bending, walking and getting in and out of a car were the most frequent forms of disability. It may be interesting to note that running on a flat surface, transferring from bed to chair, driving a car and getting in and out of the shower were infrequent causes of disability. Mean importance scores varied from 2.214 to 2.667. Physical disability occurred with high frequency, being present daily or weekly in 80 to 100% of affected patients in the majority of instances.

The prevalence of the four principal forms of physical dysfunction displayed a tendency to increase with advancing age (Fig. 2). In contrast MI scores for these four items showed little variation with age (Fig. 2)

and MI scores for the entire dimension were poorly correlated with both age ($r=0.06$) and disease duration ($r=0.08$) (Table V). MI scores for physical dysfunction were highly correlated with those for pain ($r=0.74$), moderately correlated with those for the key social and emotional items and poorly correlated with that for stiffness (Table V). No significant difference ($p=0.22$) was detected in MI scores for physical dysfunction between males and females (Table VI).

Social Dysfunction

Social function was disaggregated into seven component items which essentially captured the dimension (Table III). With the exception of restricted leisure activities (54%) social dysfunction was relatively uncommon (3 to 27%). In spite of this infrequency, the mean importance scores amongst affected individuals range from 2.154 to 2.667, and were comparable with scores for discomfort and disability on other dimensions. With the exception of dancing, social dysfunction occurred with high frequency in affected individuals (74-91%).

The prevalence of the key social disability (restricted leisure activities) showed no constant relationship to age (Fig. 3). MI scores for this item showed little variation (Fig. 3) and were poorly correlated with age ($r=-0.14$) and disease duration ($r=-0.03$) (Table V). In contrast modest correltions were noted between social dysfunction and pain, stiffness, physical and emotional dysfunction (Table V). No significant differences ($p=0.43$) were noted in MI scores for social items between males and females (Table VI).

Emotional Dysfunction

Emotional function was disaggregated into 11 component items (Table IV). Since no additional items were added, these items are considered representative of this dimension. The prevalence of emotional dysfunction ranged from 9 to 56% : anxiety, irritability, and frustration being most common. Difficulty coping with stress (17%), disturbed sense of well-being (14%) and poor self-control (9%) were infrequent problems. Mean importance scores varied from 2.111 to 2.636. With the exception of a disturbed sense of well-being and depression, emotional dysfunction occurred with high frequency in the majority of affected individuals (73-98%).

The prevalence of the key emotional disabilities showed no consistent relationship to age (Fig. 3). MI scores for these items showed little variation (Fig. 3) and were poorly correlated with both age ($r=-0.13$, -0.14, -0.20) and disease duration ($r=0.14$, -0.01, -0.01) (Table V). In contrast modest correlations were noted between key emotional items and both pain and physical function and low levels of correlation were demonstrated with stiffness (Table V). No significant differences ($p=0.51$) were noted in MI scores for emotional items between males and females (Table VI).

CONCLUSION

The assessment of pain (34) and disability (35) is a complex process which may be affected by a multiplicity of interacting biological and environmental factors (36). Not only is there significant day-to-day variability in an individual's pain sensitivity and physical performance but there is also often significant diurnal or circadian variation (37,38). Furthermore, patients with chronic musculoskeletal disease frequently show a high and unpredictable degree of individual variability in their response to therapeutic interventions (39). Thus, in attempting to measure the dimensionality of discomfort and disability in osteoarthritis, the clinical methodologist must address issues which relate to the reliability, validity and responsiveness of the measuring instrument (4). A variety of validated scales are currently

available to measure qualitative and quantitative aspects of pain (40) and various forms of physical, social and emotional disability (41). The majority of such scales are not based on musculoskeletal populations and of those which are, few have been designed to assess multidimensional symptoms in osteoarthritic patients (43).

This evaluation was specifically based on 100 patients who could fulfill musculoskeletal criteria for entry into a clinical trial of nonsteroidal anti-inflammatory drug therapy. It is evident from the data that the majority of patients surveyed experienced some form of discomfort or disability on each of the five dimensions. Pain was the most common symptom, particularly while walking, on negotiating stairs or in bed at night. Static pain whereas rated somewhat higher in importance nevertheless occurred less often. This observation is in keeping with the fact that pain at rest occurs with more severe disease and is usually preceded by pain with those activities which place joints under greater mechanical stress. It is commonly taught that joint stiffness is mild and of short duration in osteoarthritis (33). Our observations are consistent with this doctrine, at least in respect of duration, although a minority of individuals had prolonged morning stiffness. The mean importance score indicates that while often reported as "mild", for the affected individuals it is nevertheless an important source of discomfort. Difficulty negotiating stairs, standing up from a sitting position, standing still, bending, walking and getting in and out of the car occurred in the majority of patients. These locomotor disabilities were more common than those associated with less dynamic activities, e.g. putting on and taking off socks, lying in bed and sitting. However, there is one qualitative aspect of these data which should be noted. While tasks such as getting on and off the toilet cannot be avoided, others, such as heavy domestic duties and driving a car are generally avoidable and thus assume less importance than might at first be thought. Thus, amongst the patients who reported no difficulty with a given activity some were able to perform the activity without difficulty and other avoided it and therefore did not encounter any actual disability. The question regarding transfer from bed to chair proved to be a poor question since the majority of ambulant patients did not require undertaking this activity. These different conditions clearly distort the true prevalence of various forms of physical disability, nevertheless, the study provides a reasonable estimate of the proportion of patients who were concerned with any ongoing functional restriction. Since this survey was conducted for the purpose of identifying the dimensionality of pain and disability and not as an epidemiologic survey to determine the exact prevalence of each item, the principal objective was not compromised by this nuance. Restricted leisure activities were the only frequent source of social disability. The relative infrequency of other sources of social disability, clearly reflect not only avoidance of these activities due to disease but also the restriction in social activity which attends the process of ageing. Thus ageing osteoarthritics are more likely to be widowed, geographically displaced from their offspring or financially constrained.

Anxiety, irritability and frustration are common emotional responses even in healthy individuals. It is not surprising therefore that patients with osteoarthritis experience similar, albeit more intense, symptoms in this area. It is of interest that less than half the patients complained of depression and that while 67% of patients experienced pain in bed only 36% had difficulty sleeping. Furthermore, in spite of nocturnal and diurnal pain, only 17% had difficulty coping with stress and only 14% a disturbed sense of well-being. These data can likely be explained by the recurrent observation in clinical practice that patients with chronic disease accommodate to their illness. Thus, just as the expectation of cure or successful

treatment changes with time, so does the perception of pain. In addition patients find ways of minimizing the discomfort and disability which the disease causes.

We consider the questionnaire used in this survey as having face and content validity (44) by virtue of the development strategy employed, i.e. the utilization of both open-ended and closed-ended questions. Furthermore, the questionnaire was pre-tested in 15 patients with osteoarthritis and no difficulty was encountered in patient comprehension of the terminology used. Nevertheless, it might be wondered whether the questionnaire in fact probed the severity rather than the clinical importance of the patients symptoms. We do not believe this to be the case since no significant correlation was detected between either age and scores on any of the five dimensions or between disease duration and scores on any of the five dimensions. If, in fact, the questionnaire had probed severity of symptoms then given the insidiously progressive nature of the disease a time-dependent increase in symptomatology would have been expected and a moderate level of correlation detected. As anticipated the prevalence of pain and physical disability did in fact increase as a function of age although social and emotional forms of disability showed a rather inconsistent relationship. The relatively low level of correlation between age and disease duration is not entirely unexpected given the documented plateau-form age-prevalence profile of certain types of osteoarthritis (45) and the highly variable interval between the age of onset of disease and the age of onset of symptoms (46). Finally, it is only to be expected in a disease in which pain leads to disability that pain scores and disability scores would be moderately correlated.

This study highlights the multidimensionality of discomfort and disability in patients with osteoarthritis of the hip and/or knee. Health care providers and clinical investigators, therefore, require the assessment not only of the pain and of physical disability produced by this disorder but also recognition of its social and emotional consequences. Furthermore, while pain and physical disability are regarded as the most important symptoms of the disease, these data indicate that in affected individuals each symptom is regarded with similar clinical importance. Although certain physical and social activities can be avoided these data also suggest that for those disabled individuals who are still able to attempt a given activity, the importance of being able to perform the activity is similar to that of being able to perform other activities in the same and different dimensions.

At the present time no standard method exists for evaluating patients with osteoarthritis either in clinical practice or in clinical trials. Given the nature of the disease and its many differences from rheumatoid arthritis, we believe that it is timely to attempt the development of a multidimensional outcome measure for use in patients with osteoarthritis of the hip and knee. To date we have identified the dimensionality and clinical importance of a variety of pain and function items in a group of potential nonsteroidal anti-inflammatory drug trial subjects. Further work is required to assess the reliability, construct validity and responsiveness of each item and to address the issues of scaling, aggregation within- and across-dimensions and to establish the preferred method for statistical analysis. The value of using multiple items on several dimensions versus one or a few items on a restricted number of dimensions, in discriminating between an active drug and a placebo or, between two active drugs cannot be assessed from the current data but is the subject of ongoing research. Nevertheless, the items which have been identified serve as a useful battery of questions having both face and content validity which can be used to evaluate individual patients or groups of patients with osteoarthritis for descriptive purposes.

REFERENCES

1. Baum, J. A review of the psychological aspects of rheumatic disease. Semin Arthritis Rheum. 1982, 11, 352-361.
2. Cumey, H.L.F. Osteoarthritis of the hip joint and sexual activity. Ann Rheum Dis. 1970, 29, 488-493.
3. Lunghi, M.E., Miller, P.M., McQuillan, W.M. Psychological factors in osteoarthrosis of the hip. J Psychosom Res, 1978, 22, 55-63.
4. Bellamy, N., Buchanan, W.W. Outcome measurement in osteoarthritis clinical trials: The case for standardisation. Clin Rheumatol, 1984, 3, 293-303.
5. Convery, F.R., Minteer, M.A., Amiel, D., Connett, K.L. Polyarticular disability: A functional assessment. Arch Phys Med Rehabil, 1977, 58, 494-499.
6. Deniston, O.L., Jette, A. A functional status assessment instrument: Validation in an elderly population. Health Serv Res. 1980, 15, 21-34.
7. Doyle, D.V., Dieppe, P.A., Scott, J., Huskisson, E.C. An articular index for the assessment of osteoarthritis. Ann Rheum Dis. 1981, 40, 75-78.
8. Eberl, D.R., Fasching, V., Rahlafs, V., Schleyer, I., Wolfe, R. Repeatability and objectivity of various measurements in rheumatoid arthritis. Arthritis Rheum. 1976, 19, 1278-1286.
9. Empire Rheumatism Council. Multicentre controlled trial comparing cortisone acetate and acetyl salicylic acid in the long-term treatment of rheumatoid arthritis. Ann Rheum Dis. 1955, 14, 353-370.
10. Fries, J.F., Spitz, P., Kraines, R.G., Homan, H.R. Measurement of patient outcome in arthritis. Arthritis Rheum. 1980, 23, 137-145.
11. Jette, A.M. Functional capacity evaluation: An empirical approach. Arch Phys Med Rehabil. 1980, 61, 85-89.
12. Jette, A.M. Functional status index: Reliability of a chronic disease evaluative instrument. Arch Phys Med Rehabil. 1980, 62, 395-401.
13. Jette, A.M., Deniston, O.L. Interobserver reliability of a functional status instrument. J Chronic Dis. 1978, 31, 573-580.
14. Katz, S., Akpom, C.A. A measure of primary sociobiological functions. Int J Health Serv. 1976, 6, 493-507.
15. Katz, S., Downs, T.D., Cash, H.R., Grotz, R.C. Progress in development of the index of ADL. Gerontologist. 1970, 1, 20-30.
16. Katz, S., Ford, A.B., Moskiwitz, R.W., Jackson, B.A., Jaff, M.W. Studies of illness in the aged. The index of ADL: A standardised measure of biological and psychosomal function. JAMA. 1963, 914-919.
17. Keitel, W., Hoffman, H., Weber, G. Ermittlung der prozentualen Funktionsminderung der Gelenke durch einen Bewegungsfunktionstest in der Rheumatologie. Dtsch Gesundheitsw. 1971, 26, 1901-1903.
18. Lansbury, J. A method for summation of the systemic indices of rheumatoid activity. Am J Med Sci. 1956, 232, 300.
19. Lansbury, J. Report of a three-year study on the systemic and articular indexes in rheumatoid arthritis - Theoretical and clinical considerations. Arthritis Rheum. 1958, 1, 505-522.
20. Lansbury, J., Baier, H.N., McCracken, S. Statistical study of variation in systemic and articular indices. Arthritis Rheum. 1962, 5, 445-456.
21. Lansbury, J., Mueller, E.E. A numerical method for summing up total deformity, Am J Med Sci. 1958, 235, 154.
22. Lee, P., Jasani, M.K. Dick, W.C., Buchanan, W.W. Evaluation of a functional index in rheumatoid arthritis. Scand J Rheumatol. 1973, 2, 71-77.
23. Lee, P., Webb, J., Anderson, J., Buchanan, W.W. Method of assessing therapeutic potential of anti-inflammatory anti-rheumatic drugs in rheumatoid arthritis. Br Med J 1973, 2, 685-688.
24. Lequesne, M. European guidelines for clinical trials of new antirheumatic drugs. Eular Bulletin. 1980, 9 (Suppl 6), 171-175.
25. Meenan, R.F., Gertman, P.M., Mason, J.H. Measuring health status in arthritics. The arthritis impact measurement scales. Arthritis Rheum. 1980, 23, 146-152.
26. Ritchie, D.M., Boyle, J.A., McInnes, J.M., Jasani, M.K., Dalakos, T.G., Grieveson, P., Buchanan, W.W. Clinical studies with and articular index for the assessment of joint tenderness in patients with rheumatoid arthritis. 1968, Q J Med. 147, 393-406.
27. Smythe, H.A., Helewa, A., Goldsmith, C.H. Independent assessor and pooled index as techniques for measuring treatment effects in rheumatoid arthritis. J Rheumatol, 1977, 4, 144-152.
28. Steinbrocker, O., Traeger, C.H., Batterman, R.C. Therapeutic criteria in rheumatoid arthritis. JAMA, 1949, 140, 659-662.
29. Taylor D. A table for the degree of involvement in chronic arthritis. Can Med Assoc J. 1977, 36, 608-610.
30. The Co-operating Clinics Committee of the American Rheumatism Association. A seven-day variability study of 449 patients with peripheral rheumatoid arthritis. Arthritis Rheum. 1965, 8, 302-335.
31. Tugwell, P.X., Bombardier, C., Chambers, L.W. MACTAR Functional assessment questionnaire. Monograph - McMaster University - Department of Clinical Epidemiology and Biostatistics, 1981.
32. Chambers, L.W. McMaster Health Index Questionnaire (MHIQ). Monograph - McMaster University - Department of Clinical Epidemiology and Biostatistics. 1980.
33. Moskowitz, R.W. Clinical and laboratory findings in osteoarthritis. In: Arthritis and Allied Con-

ditions. Ed. McCarty, D.J. Philadelphia, Lea & Febiger, 1979, 1161-1180.
34. Melzack, R. Concepts of pain measurement. In: Pain Measurement and Assessment. Ed. Melzack, R. New York, Raven Press, 1985, 1-5.
35. International classification of impairments, disabilities and handicaps. Geneva, World Health Organization. 1980, 7-174.
36. Kremer, E.F., Block, A., Atkinson, J.H. Assessment of pain behaviour, factors that distort self-report. In: Pain Measurement and Assessment. Ed. Melzack, R. New York, Raven Press, 1985, 165-171.
37. Davis, G.C., Buchsbaum, M.S., Burney, W.E. Naloxone decreases diurnal variations in pain sensitivity and somatosensory-evoked potentials. Life Sci. 1978, 23, 1449-1459.
38. Simpson, H.W., Bellamy, N., Bohlen, J., Halberg, F. Double-blind field trial of a potential chronobiotic (Quiadon). Int J Chronobiol. 1973, 1, 287-311.
39. Bellamy, N. Between-subject variability in response to antirheumatic drugs: Implications for clinical trials. Agents Actions. 1985, (In Press).
40. Melzack, R. The McGill pain questionnaire. In: Pain Measurement and Assessment. Ed. Melzack, R. New York, Raven Press, 1985, 41-48.
41. Brook, R.H., Ware, J.E., Davies-Avery, A., Stewart, A.L., Donald, C.A., Rogers, W.K., Williams, K.N., Johnston, S.A. Overview of adult health status measures fielded in Rand's health insurance study. Med Care, 1979, 17 (Suppl 7), 1-131.
42. Deyo, R.A. Measuring functional outcomes in therapeutic trials for chronic disease. Controlled Clinical Trials. 1984, 5, 223-240.
43. Bellamy, N. The clinical evaluation of osteoarthritis in the elderly. In: Clinics in Rheumatic Diseases. Ed. Kean, W.F. Eastbourne, W.B. Saunders. 1986, (In Press).
44. Carmines, E.G., Zeller, R.A. Reliability and validity assessment. Beverley Hills, Sages Publications, 1979, 5-71.
45. Forman, M.D., Malamet, R., Kaplan, E. A survey of osteoarthritis of the knee in the elderly. J Rheumatol. 1983, 10, 282-287.
46. Bellamy, N. Osteoarthritis - An evaluative index for clinical trials. M.Sc. Thesis, Hamilton, Canada, McMaster University. 100-104.

SECTION 3 - Validation

Two major validation studies of the WOMAC Index were completed (2-4). The goal was to assess the reliability and validity of the Index, and evaluate Index responsiveness. Originally, only a single validation study, conducted within a pharmacologic randomised clinical trial (RCT), was contemplated. However, due to a lack of certainty over funding for the RCT, a trial in a total joint replacement environment was the first to be initiated. Ultimately, validation was successfully accomplished in both research settings, and undoubtedly added to the generalisibility of the observations. For the purpose of validation, responses to WOMAC questions were scaled in two different formats. The LK-scaled version allowed patients to make their responses on five-point adjectival scales (0 = none, 1 = mild, 2 = moderate, 3 = severe, 4 = extreme). In contrast, the VA-scaled version permitted responses to be made on 100 mm horizontal visual analogue scales with end markers, outside which were placed the following descriptive anchors (left end = none, right end = extreme). The validation Index used a 24-hour recall period, and instructed patients to consider the severity of their symptoms (pain, stiffness, physical function, emotional function and social function), due to OA in their hips and/or knees.

The orthopaedic study employed a quasi-experimental, one-group repeated measures design (2) in 30 OA patients undergoing total arthroplasty of the hip (n=16) or knee (n=14). Patients were evaluated the day before surgery and at six weeks, three months and six months post-operatively. In addition to LK- and VA-scaled versions of WOMAC, the following measures were administered concurrently for validation purposes: 1) Modified Doyle Index, 2) Lequesne Index, 3) Bradburn Index of Well Being, and 4) Social component of the MHIQ. Additional measures included interviewer global assessment, patient global assessment, 50 foot walk time, joint range of motion, intermalleolar straddle, and intercondylar distance.

The pharmacologic study employed a double-blind, randomised controlled trial (RCT) design (3,4) and compared two NSAIDs [Isoxicam (n= 28), Piroxicam (n = 29)] in 57 patients with OA hip (n=18) or knee (n=39). One of the major strengths of the pharmacologic validation study design was that randomisation created two groups similar in baseline characteristics and response potential. Independent evaluations of reliability, validity and responsiveness were undertaken on these two separate treatment groups, an approach which was innovative in the research environment of 1982. Patients were evaluated at enrollment and again one week later without any change in therapy in order to obtain test-retest reliability estimates at steady state. Thereafter, patients underwent a one-week NSAID-free washout period. Finally they were evaluated after two, four and six weeks of active treatment. In addition to LK- and VA-scaled versions of WOMAC, the following measures were administered concurrently for validation purposes: 1) Modified Doyle Index, 2) Lequesne Index, 3) Bradburn Index of Well Being, and 4) Social component of the MHIQ. Additional measures included interviewer global assessment, patient global assessment, 50 foot walk time, total range of motion, and intermalleolar straddle. The resulting WOMAC data were analysed by both parametric and non-parametric statistical methods to examine whether the method of analysis influenced the interpretation of the results. In the event, while the non-parametric

approach provided a more conservative estimate, both methods produced comparable results, and the method did not influence data interpretation.

The results from both studies, including both treatment groups in the pharmacologic validation study, attested to the reliability (test-retest and internal consistency), construct validity, and responsiveness of the pain, stiffness, physical function and emotional function items and subscales of the WOMAC Index. The social items did not validate well, and the social function subscale was dropped. As a consequence of decisions regarding the social function subscale, I decided to hold the emotional subscale in abeyance, and postpone its further development, pending a future redevelopment of the social function subscale. The other three subscales (pain, stiffness, physical function) performed exceptionally well, and all 24 component items were retained.

On the basis of these two validation studies, the final version of WOMAC was established. Since LK- and VA-scaled formats of the Index had been separately validated, two versions of the Index were produced: WOMAC LK 3.0 and WOMAC VA 3.0. These versions were identical with respect to item inventory and differed only in the scales on which patients responded to the component questions. The WOMAC 3.0 Index contained 24 items on three subscales: pain (5 items), stiffness (2 items), physical function (17 items), used a 24-hour time frame, and questioned patients regarding the symptom experience in their hips and/or knees. The development strategy and the results of the two validation studies, supported claims regarding the face, content and construct validity, test-retest reliability and internal consistency, and responsiveness of the WOMAC Index.

References:

2. Bellamy, N., Buchanan, W.W., Goldsmith, C.H., Campbell, J. and Stitt, L. Validation study of WOMAC: A health status instrument for measuring clinically important patient relevant outcomes following total hip or knee arthroplasty in osteoarthritis. Journal of Orthopaedic Rheumatology 1988;1:95-108.

3. Bellamy, N., Buchanan, W.W. and Grace, E. Double-blind randomized controlled trial of isoxicam vs piroxicam in elderly patients with osteoarthritis of the hip and knee. British Journal of Clinical Pharmacology 1986;22:149S-155S.

4. Bellamy, N., Buchanan, W.W., Goldsmith, C.H., Campbell, J. and Stitt, L. Validation study of WOMAC: A health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. Journal of Rheumatology 1988;15:1833-1840.

Journal of Orthopaedic Rheumatology 1, 95–108 (1988)

Validation study of WOMAC: a health status instrument for measuring clinically-important patient-relevant outcomes following total hip or knee arthroplasty in osteoarthritis

N. BELLAMY, W. W. BUCHANAN, C. H. GOLDSMITH, J. CAMPBELL and L. STITT

University of Western Ontario, London, Ontario and McMaster University, Hamilton, Ontario, Canada

The assessment of osteoarthritis patients after total joint arthroplasty of hip or knee requires the use of high performance outcome instruments of demonstrated validity, reliability, and responsiveness. We have previously developed a multi-dimensional, self-administered questionnaire with which to prove clinically-important patient-relevant outcomes. In order to validate the subscales and address issues of scaling and analysis, we have evaluated the progress of 30 osteoarthritis patients undergoing total joint arthroplasty using primary, secondary and tertiary outcome measures. The pain, stiffness and physical function subscales are valid and reliable components of a multidimensional health status instrument termed the WOMAC Osteoarthritis Index. Index responsiveness is high and the relative efficiency is greater than that of several traditional measures of surgical outcome. WOMAC complies with current clinimetric standards for evaluative instruments. In contrast to several other existing measures, its clinimetric properties are well defined, it probes clinically-important patient-relevant outcomes and it is of superior efficiency.

Keywords: outcome measurement, arthroplasty, osteoarthritis

Introduction

The main objective of evaluative research is to detect differential change in two or more treatment groups exposed to different interventions (Bellamy, 1982). The outcome instruments employed to detect change require simultaneously to be valid, reliable (Carmines and Zeller, 1979; Guion, 1974) and responsive (syn : sensitive). Of these three clinimetric properties responsiveness is of paramount importance for evaluative measures (Kirshner and Guyatt, 1985). We have recently examined the methods used in assessing the response of patients with osteoarthritis (OA) to non-steroidal anti-inflammatory agents, and have detected a high degree of variability in the outcome measurement procedures employed (Bellamy and Buchanan, 1984). A similar variability exists in the orthopaedic literature reporting the results of hip and knee arthroplasty studies.

In order to rationalize the measurement of patient-relevant outcomes in OA, we have previously probed the extent of OA symptomatology of the hip and knee by interviewing 100 consecutive consenting patients using a structured pre-tested questionnaire containing both open- and closed-ended questions probing the five dimensions of pain, stiffness, physical, social and emotional dysfunction (Bellamy and Buchanan, 1986). In that study

0951–9580/88 $03.00 + .12

Table I. Summary of item content of original test form of WOMAC*

Variable	*MIS*§	*Variable*	*MIS*§
Pain‡		*Social Function*	
1 Walking	(2.58)†	1 Leisure activities	(2.56)†
2 Stair climbing	(2.62)†	2 Community events	(2.15)†
3 Nocturnal	(2.63)†	3 Church attendance	(2.52)
4 Rest	(2.57)†	4 With spouse	(2.65)
5 Weight bearing	(2.51)†	5 With family	(2.67)
Stiffness‡		6 With friends	(2.64)
1 AMS	(2.52)†	7 With others	(2.55)
2 GEL	(2.30)†	*Emotional Function*	
Physical Function‡		1 Anxiety	(2.64)†
1 Descending stairs	(2.60)†	2 Irritability	(2.59)†
2 Ascending stairs	(2.54)†	3 Frustration	(2.44)†
3 Rising from sitting	(2.32)†	4 Depression	(2.49)
4 Standing	(2.64)†	5 Relaxation	(2.39)
5 Bending to floor	(2.51)†	6 Insomnia	(2.58)
6 Walking on flat	(2.40)†	7 Boredom	(2.62)
7 Getting in/out car	(2.26)†	8 Loneliness	(2.26)
8 Going shopping	(2.40)	9 Stress	(2.19)
9 Putting on socks	(2.38)	10 Well-being	(2.62)
10 Rising from bed	(2.37)		
11 Taking off socks	(2.37)		
12 Lying in bed	(2.36)		
13 Getting in/out bath	(2.30)		
14 Sitting	(2.54)		
15 Getting on/off toilet	(2.67)		
16 Heavy domestic duties	(2.43)		
17 Light domestic duties	(2.26)		
18 Getting on/off bus	(2.38)		

*These item numbers correspond to those in text and Table IV.
†These items were duplicated on VA scales.
‡Dimensions retained in final WOMAC instrument.
§Numbers in parentheses, represent previously published (Bellamy and Buchanan, 1986) mean importance scores (*MIS*) for each item.
(Scale: 0 = none, 1 = slight, 2 = moderate, 3 = very, 4 = extreme importance.)

we identified 50 items which characterized the clinical expression of the disorder and defined the clinical importance (MIS = mean importance score) of each. Eight of these items were subsequently discarded from further consideration for the following reasons: relevant to <10% of patients ($n = 4$), overlap with other items on same dimension ($n = 4$). In order to validate the remaining 42 items (Table I), we have recently conducted the following quasi-experimental trial (Cook and Campbell, 1979), employing a one-group repeated-measures design in patients undergoing total joint arthroplasty for OA of the hip or knee. In addition to defining the three main clinimetric properties (responsiveness, reliability and validity) of the instrument, we have also examined issues relating to scaling and statistical analysis. For convenience we have referred to this self-

administered instrument as the WOMAC (Western Ontario and McMaster Universities) Osteoarthritis Index (Bellamy, 1982). The original test instrument is illustrated in Table I.

Subjects and methods

Thirty consecutive consenting patients with primary OA of the hip or knee requiring a total joint replacement were selected for study. To be eligible patients had to fulfil the following criteria: (1) be ambulatory, (2) be unrestricted (in their functional capacity) by any associated condition, (3) not have undergone prior replacement surgery on the joint under study, (4) attend an orthopaedic clinic at one of two tertiary referral centres (University of Western Ontario or McMaster University).

Patients were assessed by trained interviewers (S.O. in London and E.G. in Hamilton) the day before surgery and at 6 weeks, 3 months and 6 months post-operatively. The same interviewer who did the initial assessment also performed all subsequent assessments on that same patient. The primary outcome measures employed were the WOMAC Osteoarthritis Index (Test Form is shown in Table I) and global assessments (on each of the five dimensions) made by both the interviewer (IGA) and patient (PGA). IGA and PGA were essential measures, since expert observers and patients themselves are regarded as key appraisers of treatment outcome (Bellamy, 1988). In order to address issues relating to scaling, patients were given (in random sequence) two versions of WOMAC to complete. Both contained identical questions but one required responses on Likert scales (Likert, 1932) while the other required responses on 10 cm horizontal visual analogue (VA) scales with terminal descriptors (Huskisson, 1982). For reasons of feasibility, while all pain and stiffness items of WOMAC were duplicated on both scales, only the first seven physical, two social, and three emotional questions (Table I) were duplicated on the VA scale. Patients completed WOMAC, IGA and PGA scales at all four assessment points. In order to test the construct validity of WOMAC, the following secondary outcome measures were concurrently applied:

(1) modified (hip and knee only) Doyle index (Doyle *et al.*, 1981)
(2) Lequesne index (Lequesne, 1980)
(3) Bradburn index of well being (Bradburn, 1969)
(4) social component of the McMaster Health Index Questionnaire (MHIQ) (Chambers, 1980).

These measures were selected as being capable of validating the five different WOMAC dimensions. Finally, four tertiary outcome measures were utilized: 50 foot walking time (*WT* – all patients), total range of movement (*ROM* – operated knees only), intermalleolar straddle (*IMS*), and intercondylar distance (*ICD*) (hip patients only). These commonly used measures of surgical outcome were selected in order to assess the Relative Efficiency (*RE*) of the final WOMAC question battery. Individual item (II) and aggregated item (AI = sum of all II for a given dimension) data were analysed for each separate WOMAC dimension using both Student's t-test (Colton, 1974a) and Wilcoxon's non-parametric test (Armitage, 1977) to assess item and dimension responsiveness and the effect of parametric versus non-parametric statistical treatment of the data. Reliability was tested on a single pre-operative administration of the instrument using Cronbach's alpha (Cronbach, 1951) and construct validity determined using Pearson's

Table II. Primary outcome measures: pre-operative means (m) and standard deviations (SD)

		Domains									
		Pain		*Stiffness*		*Physical function*		*Social function*		*Emotional function*	
Variable		*Likert**	*VA†*	*Likert**	*VA†*	*Likert**	*VA†*	*Likert**	*VA†*	*Likert**	*VA†*
WOMAC‡	*m*	11.0	248.9	4.1	100.0	41.3	378.4	9.0	108.2	11.8	121.1
	SD	4.3	117.2	1.9	57.3	14.8	162.7	7.0	59.2	8.6	78.9
PGA	*m*	2.7	63.4	2.3	59.6	2.3	54.1	1.9	48.5	1.5	33.5
	SD	0.7	24.0	1.1	22.0	1.0	25.2	1.2	28.7	1.1	29.4
IGA	*m*	2.3	56.8	2.0	48.1	2.4	58.7	1.6	35.5	1.5	33.8
	SD	0.9	24.4	0.9	24.1	0.7	13.3	0.9	21.4	0.9	25.8

*Scored on values 0–4, where 0 = none, 1 = slight, 2 = moderate, 3 = very and 4 = extreme
†0–100 mm horizontal VA scale with terminal descriptors NONE and EXTREME
‡Sum of WOMAC test questionnaire items: AI (Pain), AI (Stiffness), AI (Physical), AI (Social), AI (Emotional)

correlation coefficient (Colton, 1974b). Relative Efficiency (*RE*) was calculated using the method described by Liang and co-workers (1985) as follows:

$$RE \text{ for WOMAC versus } WT = (t_{WOMAC}/t_{WT})^2.$$

Results

Fourteen male and 16 female patients were studied. The mean age was 68.3 years (varying from 54–83) and mean disease duration (i.e. symptomatology) was 11.1 years (varying from 1–52). Fourteen patients underwent a total knee replacement and 16 received a total hip replacement. The pre-operative means and standard deviations of the study group for primary, secondary and tertiary outcome measures are illustrated in Tables II and III. One patient refused to return for assessment at 6 months and was, therefore, lost to follow-up. Therefore, the pre-operative and 6-week post-operative analyses are each based on 30 patients, while the 6-month post-operative analyses are based on 29 patients. The 3-month data, which have previously been presented in abstract form (Bellamy *et al.*, 1985), do not contribute to the interpretation of these results and have not, therefore, been reported in this paper.

PAIN

Responsiveness

On Likert scaling using Wilcoxon's test, all five pain items significantly improved by 6 weeks post-operatively ($p \leqslant 0.005$) and attained p values $\leqslant 0.001$ by 6 months post-operatively, while on VA scaling all items achieved p values of $\leqslant 0.001$ at both 6 weeks and 6 month assessments. With the IGA, PGA and AI strategies, p values of $\leqslant 0.001$ were achieved at both post-operative time points regardless of scale. When the p values derived by parametric (t-test) and non-parametric (Wilcoxon's) analysis were compared for all 64 analyses performed using II, AI, IGA and PGA strategies, there was

Table III. Secondary and tertiary outcome measures: pre-operative means (*m*) and standard deviations (SD)

	m	*SD*
Secondary		
Bradburn Total Score	1.7	4.6
Modified Doyle Total Score	2.8	1.6
Lequesne Pain Score	4.0	1.8
Lequesne Stiffness Score	1.3	0.6
Lequesne Physical Score	7.1	2.6
MHIQ Social Score	17.0	2.2
Tertiary		
WT (sec) (Hip and Knee)	26.4	12.6
IMS (cm) (Hip only)	64.6	16.5
ICD (cm) (Hip only)	45.9	9.8
ROM (°) (Operated knee only)	89.2	25.0

Table IV. Construct validity analysis: Correlation of WOMAC test items with Lequesne, modified Doyle, Bradburn and MHIQ indices

Domain		*Lequesne – pain*	*Lesquesne – stiffness*	*Lequesne – physical*	*Doyle – tenderness*	*Bradburn – emotional*	*MHIQ – social*
Pain							
Likert	1*	(0.24/0.61)	(−0.15/0.49)	(0.32/0.70)	(0.16/0.52)	(−0.18/0.39)	(−0.17/0.04)
	2†	1,2,4	2	1,3–5	2,4	3	–
	3‡	60	20	80	40	20	0
VA	1	(0.47/0.65)	(0.21/0.38)	(0.49/0.78)	(0.29/0.43)	(−0.04/0.34)	(−0.20/−0.04)
	2	1–5	1	1–5	1,4	–	–
	3	100	20	100	40	0	0
Stiffness							
Likert	1	(0.26/0.29)	(0.45/0.56)	(0.00/0.03)	(0.20/0.32)	(−0.04/0.09)	(0.27/0.28)
	2	–	AMS§,GEL¶	–	–	–	–
	3	0	100	0	0	0	0
VA	1	(0.22/0.43)	(0.32/0.35)	(0.13/0.25)	(0.07/0.12	(−0.07/0.02)	(0.05/0.17)
	2	GEL	–	–	–	–	–
	3	50	0	0	0	0	0
Physical function							
Likert	1	(0.12/0.61)	(0.00/0.36)	(0.07/0.68)	(0.14/0.64)	(−0.25/0.54)	(−0.27/0.22)
	2	5,6,8,12,14	–	1–10,12–16	2,5,7,9–11,14,15,17	11,12,14	–
	3	28	0	83	50	17	0
VA	1	(0.32/0.48)	(0.17/0.51)	(0.30/0.56)	(0.07/0.49)	(−0.20/0.32)	(−0.40/−0.03)
	2	2,4,5	1,6	1–5,7	2,7	–	5
	3	43	29	86	29	0	14

Social function							
Likert	1	(0.19/0.36)	(−0.09/0.42)	(0.14/0.58)	(0.28/0.48)	(0.24/0.51)	(−0.36/0.09)
	2	–	1	1,2	5–7	2–4	–
	3	0	14	29	43	43	0
VA	1	(0.29/0.33)	(0.13/0.39)	(0.38/0.45)	(−0.01/0.37)	(0.21/0.36)	(−0.25/0.12)
	2	–	1	1,2	–	–	–
	3	0	50	100	0	0	0
Emotional function							
Likert	1	(−0.10/0.40)	(0.24/0.35)	(−0.04/0.37)	(0.12/0.54)	(−0.02/0.59)	(−0.37/0.12)
	2	7	–	5	5,7	4–9	4
	3	10	0	10	20	60	10
VA	1	(−0.07/0.27)	(−0.29/−0.03)	(0.20/0.40)	(0.17/0.35)	(0.14/0.58)	(−0.17/−0.02)
	2	–	–	2	–	2,3	–
	3	0	0	33	0	67	0

*1 – Min/max of Pearson correlation coefficients between individual test items and comparison indices.
†2 – Test item number showing statistically significant correlation with comparison indices ($p \leq 0.05$).
‡3 – Percentage of test items showing statistically significant correlation with comparison indices.
§AMS – stiffness after first wakening in the morning.
¶GEL – stiffness after sitting, lying, or resting later in the day.
NB – Test item numbers correspond to those identified in Table I.

exact agreement (to three decimal places) in 84% of the cases while in 16% the parametric value was smaller. Correlation coefficients (Pearson's) (Colton, 1974b) between scores on Likert and VA scales were 0.88 for IGA and 0.70 for PGA.

Reliability

From Lickert-scaled responses to the five component items the reliability of the pain dimension was 0.80 pre-operatively, 0.78 at 6 weeks and 0.93 at 6 months. The corresponding values for VA-scaled responses were 0.88, 0.88 and 0.93 respectively.

Validity

Higher levels of correlation (as expressed by the correlation coefficients and the proportion of items displaying a statistically significant correlation) were noted on both Likert and VA responses between the test items and the Lequesne pain and physical function components and the Doyle Index, than between these same items and the Lequesne stiffness component, the Bradburn Index and the MHIQ social component (Table IV).

STIFFNESS

Responsiveness

On Likert scaling using Wilcoxon's test, only morning stiffness (AMS) attained statistically significant improvement by 6 weeks post-operatively ($p = 0.019$) although both AMS and gelling (stiffness occurring later in the day = GEL) significantly improved by 6 months ($p \leq 0.001$). On VA-scaled responses both items demonstrated significant improvement at both 6 weeks ($p \leq 0.001$) and 6 months ($p < 0.001$). With the IGA, PGA and AI strategies, significant improvement was detected at both post-operative time points. However, p values were smaller for these strategies (0.003, <0.001, <0.001 at 6 weeks and <0.001 for all strategies at 6 months) for VA responses than for Likert responses (0.009, 0.002, and 0.024 at 6 weeks and <0.001 for all strategies at 6 months). When the p values derived by parametric and non-parametric analyses were compared for all 40 analyses performed using II, AI, IGA and PGA strategies, there was exact agreement (to three decimal places) in 65% of the cases while in 35% the parametric value was smaller. Correlation coefficients between scores on Likert and VA scales were 0.80 for IGA and 0.56 for PGA.

Reliability

From Likert-scaled responses to the two component items, the reliability of the stiffness dimension was 0.88 pre-operatively, 0.75 at 6 weeks and 0.88 at 6 months. The corresponding values for VA-scaled responses were 0.87, 0.73, and 0.96 respectively.

Validity

Significant correlation was observed between both test items and the Lequesne stiffness component (AMS, $r = 0.45$; GEL, $r = 0.56$) and between one of the test items and the Lequesne pain component (GEL, $r = 0.43$). No significant correlation was noted between the test items and the other scales (Table IV).

PHYSICAL FUNCTION

Responsiveness

On Likert scaling using Wilcoxon's test 17 of the 18 physical function items significantly improved by 6 weeks post-operatively ($0.01 < p < 0.05$ for one item, $0.001 < p \leq 0.01$ for 6 items, and $p \leq 0.001$ for 10 items). Item 11 achieved a p value of 0.088 at 6 weeks. At 6 months, however, 17 items achieved p values of ≤ 0.001 while item eighteen attained a p value of 0.004. On VA scaling significant improvement occurred on all items, the p value being $0.01 < p < 0.05$ for one item, $0.001 < p < 0.01$ for three items, and $p \leq 0.001$ for three items. At 6 months post-operatively all seven items achieved statistically significant improvement ($p < 0.001$). With the exception of the Likert-scaled IGA response ($p = 0.133$) at 6 weeks post-operatively, the PGA and AI strategies reflected significant improvements which were mainly in the $p < 0.001$ region. When p values derived by parametric and non-parametric tests were compared for all 124 analyses performed using II, AI, IGA and PGA strategies there was exact agreement (to three decimal places) in 53% of the cases, while in 47% the parametric values were smaller. Correlation coefficients between scores on Likert and VA scales were 0.72 for IGA and 0.65 for PGA.

Reliability

From Likert-scaled responses to the 18 component items, the reliability of the physical function dimension was 0.93 pre-operatively, 0.92 at 6 weeks, and 0.97 at 6 months post-operatively. Corresponding values for the seven VA-scaled responses were 0.88, 0.91 and 0.94, respectively.

Validity

Higher levels of correlation were noted on both Likert- and VA-scaled responses between the test items and the physical component of the Lequesne Index than with these items and the Lequesne pain and stiffness components, the Doyle and Bradburn Indices, and the MIHQ social component (Table IV).

SOCIAL FUNCTION

Responsiveness

On Likert scaling using Wilcoxon's test, only the first item achieved statistically significant improvement ($p = 0.023$) at 6 weeks. By 6 months, however, five of the items had significantly improved, items six and seven attaining p values of 0.103 and 0.075, respectively. Only two items were tested on the VA scale, item one attaining a p value of <0.001 at 6 weeks, while item two was not significant ($p = 0.066$). Both items were significantly improved however by the 6 month assessment ($p < 0.001$). Although the IGA, PGA and AI strategies resulted in p values ≤ 0.004 at 6 months, the AI and PGA values at 6 weeks on Likert scaling were non-significant (0.183 and 0.211, respectively).

When the p values derived from parametric and non-parametric analyses were compared for all 60 analyses performed using II, AI, IGA and PGA strategies, there was absolute agreement (to three decimal places) in 30% of the cases, while in 57% the parametric values were smaller and in 13% the non-parametric values were smaller.

Correlation coefficients between scores on Likert and VA scales were 0.79 for IGA and 0.74 for PGA.

Reliability

From Likert-scaled responses to the seven component items, the reliability of the social function dimension was 0.86 pre-operatively, 0.88 at 6 weeks, and 0.95 at 6 months post-operatively. Corresponding values for the two VA-scaled items were 0.77, 0.88 and 0.90, respectively.

Validity

Higher levels of correlation were noted between the test items and the Doyle and Bradburn Indices (on Likert-scaled responses) than with the Lequesne stiffness or physical components (Table IV). No significant correlation was noted on the Likert scale between the test items and the Lequesne pain component or indeed with the MHIQ social component. With VA-scaled responses, higher levels of correlation were noted between test items and the Lequesne physical and stiffness component. However, no significant correlation was noted with the Lequesne pain component, the Doyle or Bradburn Indices, or again with the MHIQ social component.

EMOTIONAL FUNCTION

Responsiveness

On Likert scaling using Wilcoxon's test, none of the ten emotional function items significantly improved by 6 weeks, although by 6 months four items achieved p values ≤ 0.006, and a fifth improved to $p = 0.016$. In contrast, the three VA-scaled responses achieved p values of ≤ 0.006 at 6 weeks, and $p \leq 0.001$ at 6 months. A similar pattern emerged for Likert responses at 6 weeks on IGA ($p = 0.005$), PGA ($p = 0.170$), and AI ($p = 0.473$) strategies, with these values improving by 6 months (<0.001, 0.007 and 0.005 respectively). The corresponding VA values for IGA, PGA and AI strategies were 0.003, <0.001 and <0.001 at 6 weeks, and 0.001, 0.002 and <0.001 at 6 months. When the p values derived from parametric and non-parametric analyses were compared for all 76 analyses performed using II, AI, IGA and PGA strategies, there was exact agreement (to three decimal places) in 24% of the cases, while in 55% the parametric values were smaller, and in 21% the non-parametric values were smaller. Correlation coefficients between scores on Likert and VA scales were 0.86 for IGA and 0.78 for PGA.

Reliability

From Likert-scaled responses to the ten component questions, the reliability of the emotional function dimension was 0.89 pre-operatively, 0.92 at 6 weeks and 0.96 at 6 months post-operatively. The corresponding values for VA-scaled responses were 0.86, 0.92 and 0.97 respectively.

Validity

Higher levels of correlation were noted on both Likert- and VA-scaled responses between the test items and the Bradburn Index than with the Lequesne, Doyle or MHIQ social component (Table IV). No significant correlation was noted between test items

and the Lequesne stiffness component (on Likert responses), or between test items and VA-scaled responses on the Lequesne *pain* or *stiffness* component, the Doyle Index, or the MHIQ social component.

RELATIVE EFFICIENCY

RE was only calculated for those three dimensions showing acceptable reliability, validity and responsiveness as well as for the final (Table I) WOMAC battery (FB = AI (pain) + AI (stiffness) + AI (physical function)) (Table V). Paired t analyses of the individual hip and knee groups showed statistically significant improvements on the three dimensions and FB for each separate anatomical area. In all instances *RE* was >1, i.e. WOMAC was more efficient than any of the tertiary outcome variables (varying from 1.08–141.61).

Discussion

In developing a new health status measure we have been guided by four principles: adequate responsiveness, reliability and validity, and superior efficiency over selected existing indices.

Responsiveness

The success of total joint arthroplasty was acknowledged by both patients and interviewers, and is reflected in their PGA and IGA scores. Thirty (5 pain, 2 stiffness, 18 physical, 1 social, 4 emotional) of the original 42 WOMAC items achieved statistical significance of ≤0.006 by 6 months post-operatively. The more variable response at 6 weeks was due to some patients requiring further rehabilitation following surgery. The use of multiple analytic comparisons may result in an increase in Type I errors. Even correcting for this statistical nuance, however, and accepting a high degree of covariance amongst Index

Table V. Relative efficiency of WOMAC versus tertiary outcome variables

	*Relative efficiency**							
Tertiary Outcome Variable	*Pain*		*Stiffness*		*Physical function*		*WOMAC final battery (FB)†*	
	Likert	*VA*	*Likert*	*VA*	*Likert*	*VA*	*Likert*	*VA*
WT (Hip Only)	2.6	2.2	1.5	2.4	3.7	3.4	3.7	3.8
IMS (Hip Only)	2.6	1.7	1.1	1.3	4.9	2.3	4.7	2.3
ICD (Hip Only)	3.7	2.5	1.5	1.8	6.9	3.3	6.6	3.2
ROM (Knee Only)	134.8	79.2	104.0	86.0	132.0	67.9	141.6	88.2

*Relative Efficiency = $(t_1/t_2)^2$ eg. $(t_{pain(VA)}/t_{WT})^2 = 2.2$
†WOMAC (*FB*) = $(AI_{pain} + AI_{stiffness} + AI_{physical\ function})$

items, the p values attained were extremely good and indicative of a high level of responsiveness for these 29 WOMAC items. Likert and VA responses showed good correlation, however, since 6-month p values were often ≤ 0.001 on both scales it was difficult to distinguish between scales. The higher RE values generally achieved for Likert responses suggest that the Likert scale may be more responsive although we have a personal preference for VA scales and are now conducting additional comparative studies on scale responsiveness. At present, however, we regard both scales as legitimate for measurement purposes. Comparative analyses of non-parametric versus parametric treatment of the data suggest that while in many instances there is agreement between the two (and therefore that either analysis may be used), nevertheless, non-parametric methods provide a more conservative estimate of the response and for conceptual reasons may be regarded as the preferred analytic technique. These data show that after excluding the non-responsive social and emotional items, the remaining 30 items (5 pain, 2 stiffness, 18 physical, 1 social, 4 emotional) are of adequate responsiveness irrespective of scaling or analytic technique.

Reliability

Reliability coefficients of ≥ 0.80 are generally regarded as acceptable. Index items exceeded 0.80 on both VA and Likert scales in all except one instance (social – pre-op – VA = 0.77). These data indicate that all five WOMAC dimensions on both VA and Likert scales are of adequate reliability.

Validity

Face and content validity were conferred on Index items during an earlier investigation (Bellamy and Buchanan, 1986). For criterion validity testing coefficients ≥ 0.80 are generally regarded as acceptable. However, no gold standards currently exist against which to test criterion validity. We have, therefore, tested construct validity against other indices that probe the five Index dimensions of interest. Since these comparators are not gold standards, much lower levels of correlation are expected. In general, however, the Index items should show a statistically significant correlation with other indices probing the same dimension (convergent construct validity). Furthermore, Index items should show higher levels of correlation with indices that probe the same dimension than with indices probing other dimensions (divergent construct validity). These criteria were fulfilled by the pain, stiffness and physical function components of the Index. Since physical disability is often secondary to pain, it is not surprising that these two dimensions are often associated. In contrast, the social component failed to correlate with the MHIQ social component, and although some items were reliable and responsive, this dimension was excluded from the Index. Moreover, in spite of the emotional component fulfilling construct validity criteria, and some items being reliable and responsive, we have elected to withdraw the component pending a re-evaluation of the social dimension. The final Index, therefore, utilizes the pain (5 items), stiffness (2 items) and physical function (18 items) subscales only (Table I).

Relative efficiency

To be useful, a new health status measure should offer advantages over existing indices. In this respect, WOMAC offers two advantages. First, WOMAC and its subscales offer superior efficiency over the traditional unidimensional outcome measures selected, both

in assessing the results of total hip arthroplasty and also the results of total knee arthroplasty. Such a measure has potential for reducing sample size requirements for clinical trials using WOMAC as the primary outcome measure (cf. these other measures). It should be noted, however, that there are several other measures (some of which are also multi-dimensional) against which WOMAC has yet to be compared. Secondly, many of the traditional unidimensional measures lack patient relevance. In contrast, WOMAC probes patient-relevant outcomes, the clinical importance (Table I) of which have been documented and reported previously (Bellamy and Buchanan, 1986).

WOMAC is a reliable, valid and responsive multi-dimensional, self-administered questionnaire having greater efficiency than several traditional unidimensional measures. We are now conducting further studies on aggregating scores across different dimensions, on the relative responsiveness of Likert and VA scales, and the relative efficiency of WOMAC against several other multi-dimensional instruments used in assessing the results of total hip and knee joint arthroplasty.

Acknowledgements

We wish to thank the following for their contribution to the development of WOMAC; H. W. Bailey, S. Bailey, S. O'Leary, E. Grace, and E. Duku. This project was supported in part by a grant from The Arthritis Society of Canada.

References

Armitage, P. (1977) Distribution-Free Methods. In *Statistical Methods in Medical Research*, Chap. 13, pp. 394–407. New York: John Wiley and Sons.

Bellamy, N. (1982) Osteoarthritis – An Evaluative Index for Clinical Trials. M.Sc. Thesis. McMaster University, Hamilton, Canada.

Bellamy, N. (1988) Methods of clinical assessment of anti-rheumatic drugs. In *Clinical Rheumatology: Anti-rheumatic drugs* (edited by P. Brooks). London: Balliere Tindall. (In Press.)

Bellamy, N. and Buchanan, W.W. (1984) Outcome measurement in osteoarthritis clinical trials: The case for standardisation, *Clinical Rheumatology* 3(3), 293–305.

Bellamy, N., O'Leary, S., Grace, E. and Buchanan, W.W. (1985) Osteoarthritis: An assessment of signal versus global strategies in outcome measurement. XVIth International Congress of Rheumatology, Sydney, Australia, Abstract Book, p. 50. International League Against Rheumatism.

Bellamy, N., O'Leary, S., Grace, E., and Buchanan, W.W. (1985) Osteoarthritis: An assessment of signal versus global strategies in outcome measurement. XVIth International Congress of Rheumatology, Sydney, Australia, Abstract Book, p. 50.

Bradburn, N.M. (1969) *The Structure of Psychological Well-Being*. Chicago: Aldine Publishing Co.

Carmines, E.G. and Zeller, R.A. (1979) *Reliability and Validity Assessment*. Beverly Hills and London: Sage Publications.

Chambers, L.W. (1980) *McMaster Health Index Questionnaire (MHIQ)*. Hamilton, Canada: McMaster University, Department of Clinical Epidemiology and Biostatistics.

Colton, T. (1974a) Inference on Means. In *Statistics in Medicine*, Chap. 4, pp. 99–150. Boston: Little Brown and Co.

Colton, T. (1974b) Regression and Correlation. In *Statistics in Medicine*, Chap. 6, pp. 189–217. Boston: Little Brown and Co.

Cook, T.D. and Campbell, D.T. (1979) Quasi-experiments: Nonequivalent control group designs. In *Quasi-Experimentation: Design and Analysis Issues for Field Settings*, Chap. 3, pp. 95–146. Boston: Houghton Mifflin Company.

Cronbach, L.J. (1951) Coefficient alpha and the internal structure of tests. *Psychometrika* **16**, 297–334.

Doyle, D.V., Dieppe, P.A., Scott, J. and Huskisson, E.C. (1981) An articular index for the assessment of osteoarthritis, *Annals of the Rheumatic Diseases* **40**, 75–8.

Guion, R.M. (1974) Reliability and Validity. In *Standards for Educational and Psychological Tests*, pp. 25–55. Washington, DC: American Psychological Association Inc.

Huskisson, E.C. (1982) Measurement of pain, *Journal of Rheumatology* 9, 768–9.

Kirshner, B. and Guyatt, G. (1985) A methodological framework for assessing health indices, *Journal of Chronic Diseases* **38(1)**, 27–36.

Lequesne, M. (1980) European guidelines for clinical trials of new antirheumatic drugs, *Eular Bulletin* **9(Supplement 6)**, 171–5.

Liang, M.H., Larson, M.G., Cullen, K.E. and Schwartz, J.A. (1985) Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research, *Arthritis and Rheumatism* **28(5)**, 542–7.

Likert, R. (1932) A technique for the measurement of attitudes, *Archives of Psychology* **140**, 44–60.

Accepted for publication 4 April 1988

Br. J. clin. Pharmac. (1986), 22, 149S–155S

Double-blind randomized controlled trial of isoxicam *vs* piroxicam in elderly patients with osteoarthritis of the hip and knee

N. BELLAMY[1,2], W. W. BUCHANAN[2] & E. GRACE[2]
[1]University of Western Ontario, London, Canada and [2]McMaster University, Hamilton, Ontario, Canada

1 Fifty-seven elderly patients with primary osteoarthritis of the hip and knee were entered into a double-blind, randomized, controlled parallel group trial to compare the efficacy and tolerability of isoxicam (maximum = 200 mg day^{-1}) and piroxicam (maximum = 20 mg day^{-1}).

2 Clinical assessments were made following a 1 week NSAID-free washout period and at biweekly intervals during the next 6 weeks of active treatment.

3 The majority of patients in both groups experienced a clinically important and statistically significant therapeutic response.

4 No statistically significant between-group differences were noted with respect to drug efficacy.

5 One patient was withdrawn from the piroxicam group because of lack of effect, but there were no such withdrawals from the isoxicam group.

6 Five patients were withdrawn from the piroxicam group because of adverse reactions compared to only one withdrawal from the isoxicam group.

7 This study indicates that isoxicam is an efficacious and well-tolerated once-daily NSAID for elderly patients with osteoarthritis.

Keywords osteoarthritis isoxicam piroxicam elderly

Introduction

Osteoarthritis is primarily a disease of the elderly (Kellgren & Lawrence, 1958). Although many individuals have subclinical or pre-radiographic disease, the prevalence of symptomatic disease shows a positive correlation with chronologic age. Although primary osteoarthritis may affect the distal and proximal interphalangeal joints and the first carpometacarpal joints of the hands as well as the apophyseal joints of the axial skeleton and the first metatarsophalangeal joint of the toe, it is involvement of the hip and knee which results in the greatest discomfort and locomotor disability. Since there are currently no disease-modifying anti-rheumatic drugs with which to treat osteoarthritis, the principal objective of therapy is to relieve pain and improve function. Thus the non-steroidal, analgesic anti-inflammatory drug remain the treatment of choice for patients with this disorder. The response to any given agent, varies between individuals and has in recent years resulted in the development of a number of new non-steroidal anti-inflammatory drugs (NSAIDs) (Rosenbloom *et al.*, 1985). The elderly in particular require an agent which is not only efficacious and well tolerated but also can

be taken according to a simple dosing regimen. There has been a steadily growing interest in studying therapeutic responsiveness and tolerability in elderly subjects since they represent a large proportion of the population with chronic musculoskeletal disease. Furthermore, they have a limited potential for response and demonstrate an increased propensity for developing adverse reactions to certain classes of drugs. In the report which follows, the efficacy and tolerability of isoxicam, a new NSAID, have been compared to those of piroxicam in a group of elderly patients with osteoarthritis of the hip and knee. Isoxicam has the chemical formula 4-hydroxy-2-methyl-*N*-(5-methyl-3-isoxolyl)-2H-1, 2-benzothiazine-3-carboxamide 1, 1-dioxide and is a benzothiazine derivative of the oxicam class of drugs (Yakatan, 1982). In addition to its comparative objectives, this trial was used to validate a new multidimensional outcome measure for clinical trials in osteoarthritis (Bellamy, 1982). The results of the validation study form the basis of a separate report and are not discussed in this paper.

Methods

The trial was conducted in two centres and employed a double-blind, randomized, controlled parallel design. Elderly patients with primary osteoarthritis of the hip and/or knee were enrolled in the study programme if they were 55–85 years of age and showed radiographic evidence of osteoarthritis. They were excluded if they had undergone prior replacement surgery on the joint of interest or had secondary osteoarthritis, active peptic ulceration or prior gastrointestinal bleeding. Other exclusion criteria included cardiorespiratory insufficiency, significant disease of any other major organ system, a BUN greater than 30 mg 100 ml^{-1}, SGOT greater than 50 units ml^{-1}, allergy to aspirin or other NSAIDs, concurrent anticoagulant therapy, or recent systemic or intra-articular corticosteroid therapy. Following enrollment, patients were assessed and randomized to receive either isoxicam or piroxicam according to a randomization code developed by an independent statistician. Thereafter patients underwent a 1-week NSAID-free washout period during which time analgesia with paracetamol (325 mg tablets) was permitted but monitored.

Following this washout period, baseline assessments were made and active treatment started with either isoxicam, 100 mg day^{-1}, or piroxicam, 10 mg day^{-1} (Level 1). Each medication was identical in appearance and presented in capsule form. Patients were instructed to take the medication once daily with their breakfast and were advised against the concomitant use of salicylate-containing compounds during the study period. Patients were reassessed 2 weeks later and a decision made to continue at this treatment level or, if the response had been suboptimal, to increase the dosage to either isoxicam, 200 mg day^{-1}, or piroxicam, 20 mg day^{-1} (Level 2).

Patients were again assessed 2 weeks later and a final assessment was made 2 weeks after that. The total period of active treatment was 6 weeks. The following outcome measures were employed, assessments being made at each of the four assessment points:

1. Night pain (four-point Likert scale (Nunnally, 1967))
2. Pain on walking (four-point Likert scale)
3. Degree of starting pain (four-point Likert scale)
4. Pain on joint motion (four-point Likert scale)
5. Joint tenderness (modified Doyle index) (Doyle *et al.*, 1981)
6. Knee total range of movement (degrees)
7. Intermalleolar straddle (cm)
8. Visual analogue scale of pain (vertical, 21-compartment, terminal descriptors only)
9. Physician overall assessment (five-point Likert scale)
10. Patient overall assessment (five-point Likert scale)
11. Walking time (50 feet)
12. Laboratory values: CBC, SMA-12, urinalysis.

Many patients had multijoint involvement but for each patient a single (either a hip or a knee) joint was identified as being more severely affected and was selected as the target joint for outcome measures numbers 1–7. In addition, the investigators recorded any adverse reactions to drug therapy and concurrent illness. At the final assessment (visit 4), both patients and physicians compared the efficacy and tolerability of the study drugs to both the drug-free washout period and their pre-study anti-inflammatory medication. During the course of the trial patients were allowed to take concomitant analgesia with paracetamol (325 mg tablets) although patients were instructed to use the tablets sparingly and their requirements were monitored. Compliance to the study medications was measured by both direct report and by pill count.

The analysis was conducted using an intention-to-treat philosophy rather than an explicative approach (Sackett & Gent, 1979). Thus withdrawals were accounted for in the analysis rather than being excluded from consideration. The following statistical tests were employed: chi square (Fleiss, 1981), Fischer's exact (Fleiss, 1981), unpaired two-tailed Student's *t*-test (Armitage, 1971), Mann Whitney U (for visual analogue and Likert-type scales) (Conover, 1980), and log rank chi square (for life table analysis (Friedman *et al.*, 1981).

Results

Of the 57 patients enrolled in the study, 28 received isoxicam and 29 received piroxicam. The mean age was exactly the same in each group (66.5 years) and the age ranges were similar (Table 1). Mean disease duration was slightly greater for isoxicam (9.3 years) than for piroxicam (8.7 years), but this difference was not statistically significant. Furthermore no significant differences were noted in the proportion of hips and knees represented in the two treatment groups, although in general the knee joint was frequently the more severely affected of the two joints. Males and females were represented in similar proportions although there was a slight female predominance in the piroxicam group (Table 1). Post-randomization group comparability was assessed at the end of the washout period. No statistically significant between-group differences were detected for any demographic or disease variable (Table 2).

In order to evaluate the response achieved at the initial level of therapy (i.e., isoxicam, 100 mg day^{-1}, or piroxicam, 10 mg day^{-1}) data were analyzed at visit 2 and compared with baseline values (Table 3). Significant improvements were noted in pain on joint movement, visual analogue scale for pain, and the physician's overall assessment in both groups. In the isoxicam group, significant improvement was also detected in pain on walking and patient overall assessment, while in the piroxicam group significant improvement was noted in night pain and joint tenderness. Overall, both groups showed improvement at the initial level of therapy, although only six isoxicam and four piroxicam patients remained at this level (i.e., isoxicam, 100 mg day^{-1}, or piroxicam, 10 mg day^{-1}). A between-group comparison made at visit 3 showed no statistically significant difference other than in relief of pain on walking, which favoured the isoxicam group.

In order to assess response to the optimal dose, i.e., after any necessary dose titration, comparisons were made between visit 4 and baseline. Statistically significant improvement over baseline (Table 4) was noted in both treatment groups with respect to night pain, pain on walking, starting pain, pain on movement, joint tenderness, visual analogue scale of pain, overall assessment by physician, and overall assessment by patient. Although joint swelling in the knee declined in both groups (Table 2) a statistically significant response was achieved only in the isoxicam group (Table 4). Knee range of movement and internalleolar straddle increased and the walking time decreased in both groups (Table 2) although none of these measures reached statistical significance in either group (Table 4). Between-group comparisons were made for all disease variables at visit 4 (i.e., after 6 weeks of active treatment); no significant differences were detected, indicating that the therapeutic response was similar in the two treatment groups (Table 4). One patient was withdrawn from the piroxicam group because of inefficacy; there were no such withdrawals from the isoxicam group. Compliance levels for study medications were exceptionally high both when measured by direct report and by pill count. Data collection on concomitant analgesia was incomplete and while no apparent between-group differences were noted no formal analysis could be performed.

Twelve adverse reactions in six patients were reported on isoxicam *vs* 24 adverse reactions in 12 patients on piroxicam ($P = 0.105$) (Table 5). Fluid retention, gastrointestinal upset and neurological symptoms accounted for the majority of adverse reactions. Reactions were graded by the investigators as being either mild, moderate or severe. Although five severe adverse

Table 1 Post-randomization – pre-intervention comparison of treatment groups

Variable	*Isoxicam (n = 28)*	*Piroxicam* (n = *29)*	P *value*
Age (range) (years)	66.5 (55–80)	66.5 yrs. (55–82)	NS
Disease duration (years)	9.3 (1–26)	8.7 yrs. (1–30)	NS
Sex	14 male 14 female	12 male 17 female	NS
Most severely affected joint	7 hip 21 knee	11 hip 18 knee	NS

Table 2 Mean values for outcome variables at baseline and visits 2, 3 and 4. S.d. given for baseline values

		Baseline	*Visit 2*	*Visit 3*	*Visit 4*
Night pain	(I)	1.68 ± (0.72)	1.26	0.93	0.63
	(P)	1.83 ± (1.00)	1.25	1.19	0.77
Pain on walking	(I)	2.00 ± (0.82)	1.30	1.30	1.00
	(P)	2.14 ± (0.92)	1.86	1.36	1.05
Standing pain	(I)	1.39 ± (0.99)	1.04	0.78	0.65
	(P)	1.79 ± (1.11)	1.54	1.04	0.82
Pain on movement	(I)	1.57 ± (1.00)	1.04	0.59	0.52
	(P)	2.07 ± (0.75)	1.36	1.23	0.59
Joint tenderness	(I)	1.50 ± (0.92)	1.00	0.62	0.56
	(P)	1.89 ± (0.82)	1.04	0.91	0.55
Joint swelling (knee)	(I)	0.52 ± (0.75)	0.30	0.35	0.15
	(P)	0.94 ± (1.03)	0.59	0.65	0.38
Range of movement (knee) (°)	(I)	106.48 ± (22.17)	113.75	116.55	115.55
	(P)	102.44 ± (15.47)	106.24	106.88	109.50
Intermalleolar straddle (hip) (mm)	(I)	71.14 ± (27.01)	77.86	85.14	91.29
	(P)	60.00 ± (19.10)	60.20	67.63	74.17
Analogue scale of pain (mm)	(I)	13.71 ± (5.33)	8.85	7.41	6.19
	(P)	14.17 ± (5.34)	10.79	8.68	8.68
Overall assessment (MD)	(I)	3.44 ± (0.85)	2.67	2.46	2.15
	(P)	3.52 ± (0.87)	2.96	2.56	2.27
Overall assessment (patient)	(I)	3.48 ± (0.80)	2.70	2.54	2.26
	(P)	3.52 ± (0.83)	3.00	2.68	2.32
Walking time (s)	(I)	17.14 ± (8.26)	15.27	14.41	14.72
	(P)	15.25 ± (6.02)	15.96	15.90	14.39

I isoxicam, P piroxicam

reactions were noted on piroxicam, none of the reactions observed on isoxicam (Table 5) were graded as severe ($P = 0.03$). Furthermore, five patients were withdrawn from piroxicam due to adverse reactions compared to only one withdrawal from isoxicam (a moderate adverse reaction) ($P = 0.105$).

Thus, the total number of patients withdrawn due to either inefficacy or intolerance was six in the piroxicam group and one in the isoxicam group. The ability of patients to stay in treatment was compared using a life table approach (Friedman *et al.*, 1981). Although the between-group differences were not statistically significant ($P > 0.05$), it can be seen from Figure 1 that apart from one early withdrawal on isoxicam, the remaining patients stayed on treatment. In contrast, patients dropped out of piroxicam

Table 3 Statistical analysis of outcome variables

	Baseline *Isoxicam* vs *piroxicam*	*Visit 2* vs *baseline isoxicam*	*Level 1* *Visit 2* vs *baseline piroxicam*	*Visit 2* *Isoxicam* vs *piroxicam*
Night pain	0.36	0.06	0.04*	0.99
Pain on walking	0.37	< 0.01*	0.07	0.02*
Standing pain	0.17	0.21	0.34	0.07
Pain on movement	> 0.05	0.05*	< 0.01*	0.16
Joint tenderness	0.09	0.06	< 0.01*	0.99
Joint swelling (knee)	0.21	0.32	0.36	0.16
Range of movement (knee)	0.52	0.22	0.43	0.10
Intermalleolar straddle (hips)	0.33	0.64	0.98	0.15
Analogue scale of pain	0.66	< 0.01*	0.02*	0.24
Overall assessment (MD)	0.65	< 0.01*	0.04*	0.21
Overall assessment (patient)	0.90	< 0.01*	0.07	0.28
Walking time	0.96	0.40	0.42	0.72

* $P \leq 0.05$

Table 4 Statistical analysis of outcome variables

	Baseline		*Level 1 + 2*	
	Isoxicam vs *piroxicam*	*Visit 4* vs *baseline isoxicam*	*Visit 4* vs *baseline piroxicam*	*Visit 4 Isoxicam* vs *piroxicam*
Night pain	0.36	< 0.01*	< 0.01*	0.51
Pain on walking	0.37	< 0.01*	< 0.01*	0.75
Standing pain	0.17	< 0.01*	< 0.01*	0.46
Pain on movement	> 0.05	< 0.01*	< 0.01*	0.49
Joint tenderness	0.09	< 0.01*	< 0.01*	0.91
Joint swelling (knee)	0.21	0.04*	0.10	0.13
Range of movement (knee)	0.52	0.14	0.14	0.20
Intermalleolar straddle (hips)	0.33	0.16	0.18	0.19
Analogue scale of pain	0.66	< 0.01*	< 0.01*	0.18
Overall assessment (MD)	0.65	< 0.01*	< 0.01*	0.66
Overall assessment (patient)	0.90	< 0.01*	< 0.01*	0.99
Walking time	0.96	0.24	0.07	0.84

* $P \leq 0.05$

therapy at intervals and as late as the thirty-eighth day of treatment.

An end-of-study comparison between the active treatment phase and the washout period, based on patient evaluations, indicated that 93% of isoxicam patients had improved *vs* 69% of piroxicam patients. Physician-based estimates for these same comparisons were 93% and 75% respectively. In the isoxicam group, 73% of patients and 72% of physicians considered the study drug better than the patient's pre-study medication. In the piroxicam group, 61% of patients and 73% of physicians considered the study drug better than patient's pre-study medication. Eighty-nine percent of patients and 96% of physicians rated isoxicam to have been as well or better tolerated than the pre-study medication.

Table 5 Adverse reactions reported by the two treatment groups

Reactions	*Isoxicam* (n = 28)	*Piroxicam* (n = 29)	
Abdominal cramps	1	1	
Ankle swelling	1	5	
Asthenia	0	1	
Diarrhoea	0	1	
Dizziness	0	2	
Drowsiness	1	2	
Dry mouth	2	1	
Epigastric pain	0	2	
Hot flushes	1	1	
Metallic taste	1	1	
Nausea and vomiting	0	1	
Oral ulcer	0	1	
Nightmare	1	0	
Shortness of breath	1	3	
Skin rash	0	1	
Swollen ear, Difficulty swallowing	0	1	
Tinnitus	2	0	
Tingling of tongue	1	0	
Total number of adverse reactions (ADR) +	12	24	
Total number of patients experiencing ADR	6	12	$P = 0.105$
Total number of patients with severe ADR	0	5	$P = 0.03$*
Total number of patients withdrawn due to ADR++	1	5	$P = 0.105$
Total number of patients withdrawn due to either inefficacy or ADR	1	6	$P = 0.056$

+ Some patients had more than one adverse reaction.
++ One patient withdrawn due to a moderate adverse reaction.
* $P \leq 0.05$

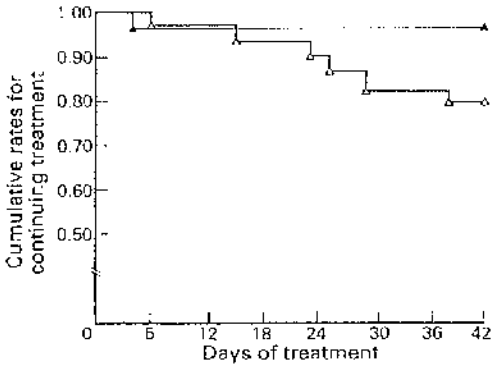


Figure 1 Log rank Chi square comparison of cumulative rates (for continuing treatment (▲ isoxicam, △ piroxicam). Note break in vertical scale.

Derangements of haematological and biochemical variables were infrequent and no extreme abnormalities were noted. The haemoglobin fell below the lower limit of the normal range in three isoxicam and three piroxicam treated patients. Elevation of the serum creatinine was observed in one isoxicam and two piroxicam treated patients, elevation of the SGOT in two piroxicam treated patients, and elevation of the alkaline phosphatase in two isoxicam patients.

Discussion

The elderly represent a subset of the general population in whom degenerative forms of arthritis are common, concurrent illness not infrequent, and multiple drug therapy is prevalent. The study of the elderly is particularly important since individuals in this age group are frequent recipients of drugs which have been evaluated largely in younger subjects. Thus, given the relative frequency of osteoarthritis and its tendency to affect the middle-aged and elderly, this trial provides important efficacy and tolerability data on isoxicam in this subset of the population with musculoskeletal-disease. Although the titration strategy for dose adjustment does not allow simple dose-by-dose comparison (because of the small residual sample sizes), it does nevertheless allow isoxicam and piroxicam to be compared under simulated practice conditions.

The results of this study can be generalized to similar elderly patients fulfilling the defined inclusion and exclusion criteria and those who are treated under the described titration strategy. It is traditional in comparative studies of this type to exclude high-risk patients, particularly those with severe disease of any major organ system or prior gastrointestinal bleeding. Accordingly, the results of this clinical trial cannot necessarily be extrapolated to patients with serious concomitant disease.

The result of this study indicate that isoxicam is an efficacious non-steroidal anti-inflammatory agent which significantly reduces pain, swelling and joint tenderness in elderly patients with osteoarthritis. Although some patients demonstrate a response to 100 mg day^{-1} of isoxicam, the majority require a dosage of 200 mg day^{-1}. The beneficial effects of isoxicam are in general similar to those of the drug piroxicam.

Three measures of physical function (knee range of movement, intermalleolar straddle, and 50 foot walking time) failed to improve on either agent. Two factors may explain this failure in response. First, since range of knee movement was analyzed only for patients in whom the knee was the most severely affected joint and intermalleolar straddle only analyzed for those in whom the hip was the most severely affected joint, the sample sizes employed in the analysis were small and therefore the statistical power was low. Second, both these measures are attended by significant inter-observer variation which may have reduced the chance of detecting a statistically significant improvement. Response failure on the 50-foot walking time was not entirely unexpected as this outcome measure has previously displayed poor performance characteristics in clinical trials in osteoarthritis (Bellamy & Buchanan, 1984). Since pain is often the limiting factor in degenerative forms of arthritis, it is not surprising that relief of pain might enhance the ease of performing a particular task without necessarily improving the speed at which it can be accomplished. Although it is purely speculative, this may provide an explanation for the poor performance of the 50-foot walking time. An alternate explanation is that the difference between a healthy and a diseased elderly patient with osteoarthritis in respect of the ability to perform the walking time is less than between a healthy young individual and a young arthritic patient and, therefore, the response potential is more restricted in the elderly.

It is important that a new non-steroidal anti-inflammatory agent for elderly patients be not only efficacious but also well tolerated. The discontinuation of only one isoxicam patient due to a moderate adverse reaction compared to five piroxicam patients suggests that, in general, isoxicam may be better tolerated by the elderly

than piroxicam. Although the life table approach is infrequently used in the analysis of drug trials it has the ability to discriminate between different patterns of withdrawal from treatment. Whilst between-group differences were not statistically significant, nevertheless the ability of patients to stay on treatment with isoxicam was notable. These data suggest that isoxicam may be a useful drug in the treatment of osteoarthritis in the elderly.

References

Armitage, P. (1971). *Statistical methods in medical research*, pp 119–126. Boston: Blackwell Scientific Publications.

Bellamy, N. (1982). *Osteoarthritis — an evaluative index for clinical trials*. McMaster University Thesis.

Bellamy, N. & Buchanan, W. W. (1984). Outcome measurement in osteoarthritis clinical trials: a case for standardisation. *Clin. Rheum.*, **3**, 293–304.

Conover, W. J. (1980). *Practical non-parametric statistics*, pp 216–227 Toronto: John Wiley & Sons.

Doyle, D. V., Dieppe, P. A., Scott, J. & Huskisson, E. C. (1981). An articular index for the assessment of osteoarthritis. *Ann. Rheum. Dis.*, **40**, 75–78.

Fleiss, J. L. (1981). *Statistical methods for ratio proportion*, pp 19–29, 24–26. Toronto: John Wiley & Sons.

Friedman, L. M., Furberg, C. D. & DeMets, D. L. (1981). *Fundamentals of clinical trials*, pp 180–193. Boston: John Wright PSG, Inc.

Kellgren, J. H. & Lawrence, J. S. (1958). Osteoarthritis and disc degeneration in an urban population. *Ann. Rheum. Dis.*, **17**, 388–397.

Nunnally, J. C. (1967). *Psychometric theory*, pp. 521–523. New York: McGraw Hill.

Rosenbloom, D., Brooks, P. M., Bellamy, N. & Buchanan, W. W. (1985). Anti-rheumatic drug therapy: a critical review of clinical trials. *Praeger Scientific*, (in press).

Sackett, D. L. & Gent, M. (1979). Controversy in counting and attributing events in clinical trials. *New Engl. J. Med.*, **301**, 1410–1412.

Yakatan, G. J. (1982). Pharmacology and pharmacokinetics of isoxicam. *Sem. Arth. Rheum.*, **12**, 154–159.

Validation Study of WOMAC: A Health Status Instrument for Measuring Clinically Important Patient Relevant Outcomes to Antirheumatic Drug Therapy in Patients with Osteoarthritis of the Hip or Knee

NICHOLAS BELLAMY, W. WATSON BUCHANAN, CHARLES H. GOLDSMITH, JANE CAMPBELL, and LARRY W. STITT

Abstract. **Within the context of a double blind randomized controlled parallel trial of 2 nonsteroidal antiinflammatory drugs, we validated WOMAC, a new multidimensional, self-administered health status instrument for patients with osteoarthritis of the hip or knee. The pain, stiffness and physical function subscales fulfil conventional criteria for face, content and construct validity, reliability, responsiveness and relative efficiency. WOMAC is a disease-specific purpose built high performance instrument for evaluative research in osteoarthritis clinical trials.** *(J Rheumatol 1988;15:1833-1840)*

Key Indexing Terms:
WOMAC OSTEOARTHRITIS VALIDITY
HEALTH STATUS INSTRUMENT RELIABILITY RESPONSIVENESS

We reported on the inadequacy of outcome measurement procedures in osteoarthritis (OA) trials of nonsteroidal antiinflammatory (NSAID) drugs[1]. In an attempt to rationalize measurement in OA, we first probed the symptomatology of hip and knee OA by interviewing 100 patients with OA, and identified 41 items on 5 dimensions (Table 1) which characterize the disorder[2,3]. To validate these items, we have now conducted a double blind, randomized, controlled parallel design trial of isoxicam versus piroxicam in elderly patients with primary OA of the hip or knee. Our goal was to assess the reliability, construct validity and responsiveness of the 41 items previously mentioned. The 2 separate arms of the trial served as independent tests of item responsiveness. To circumvent the problem of simultaneously attempting to assess the performance of a new health status instrument as well as a new antirheumatic compound, we used other reported[4] outcome measures to evaluate the comparative efficacy and tolerability of isoxicam and piroxicam. Thus, using this innovative approach, we defined the clinimetric properties of a health status instrument termed "WOMAC" (the Western Ontario and McMaster Universities Osteoarthritis Index), within the context of a traditional clinical trial.

MATERIALS AND METHODS

Fifty-seven patients with symptomatic OA of the hip or knee requiring NSAID therapy were entered in the study. To be eligible patients had to be 55–85 years of age, have definite radiographic evidence of primary OA in the hip or knee, and fulfil defined inclusion and exclusion criteria[4].

Patients were assessed at enrolment (Visit 1), and again one week later without any change in therapy (Visit 2). Thereafter, patients underwent a one-week NSAID-free washout period and were then reassessed (Visit 3). Finally, patients were evaluated following 2 (Visit 4), 4 (Visit 5), and 6 (Visit 6) weeks of active treatment. The initial drug dosage was piroxicam 10 mg OD or isoxicam 100 mg OD, this being increased at Visit 4 to 20 mg OD or 200 mg OD, respectively, in patients failing to respond to the lower dosage[4].

The primary outcome measures employed were the WOMAC OA Index (Test Form, Table 1) and 2 forms of global assessment. WOMAC was self-administered while the global assessments (on each of the 5 dimensions) were made both by trained interviewers (interviewer global assessment) and study patients (patient global assessment). To address issues relating to scaling, patients were given (in random sequence) 2 versions of WOMAC to complete. Both contained identical questions but one required responses on 5-point (none, slight, moderate, severe, extreme) Likert scales[5] while the other required responses on 10 cm horizontal visual analogue scales (VAS) with terminal descriptors[6]. Individual item scores were determined by reading the patient's response to each question. Aggregate scores for each dimension were determined by summing the component item scores for each dimension. The WOMAC final battery was determined by summing the aggregate scores for the pain, stiffness and physical function dimensions. For reasons of feasibility, only the first 3 pain and 7 physical questions were duplicated on both scales (Table 1). Similarly, only the question pertaining to severity of morning stiffness, the 1st, 2nd and 6th social, and the 1st, 2nd, 4th and 5th emotional questions were duplicated on the VAS

From the University of Western Ontario, London, and McMaster University, Hamilton, ON, Canada.

Supported in part by a grant from the Arthritis Society of Canada and in part by a grant from Parke-Davis Canada Inc.

N. Bellamy, MD, MSc, FRCP(Edin, Glas, and C), FACP, Associate Professor of Medicine, Epidemiology and Biostatistics, University of Western Ontario; W.W. Buchanan, MD, FRCP(Edin, Glas, and C), Professor of Medicine; C.H. Goldsmith, PhD, Professor of Clinical Epidemiology and Biostatistics, McMaster University; J. Campbell, BA, Research Assistant, University of Western Ontario; L.W. Stitt, MSc, Statistical Consultant/Analyst and Lecturer, Department of Statistical and Actuarial Sciences, University of Western Ontario.

Address requests for reprints to Dr. N. Bellamy, Ste. 402A, W.T., Victoria Hospital, 800 Commissioners Rd. E., London, ON, Canada, N6A 4G5.

Submitted August 19, 1988 revision accepted September 21, 1988.

scale (Table 1). As with WOMAC, the interviewer and patient global assessment scores on single questions which separately probed the overall status of the patient on each of the 5 dimensions were made on both Likert and VAS scales. Patients completed WOMAC, interviewer and patient global assessments at all 6 visits. To test the construct validity of WOMAC, the following secondary outcome measures were concurrently applied: (1) joint tenderness (modified Doyle Index [hip and knee only])[7], (2) Lequesne Index[8], (3) Bradburn Index of Well Being[9], and (4) social component of the McMaster Health Index Questionnaire (MHIQ)[10]. These measures were selected as being capable of validating the 5 different WOMAC dimensions i.e., Pain (Doyle, Lequesne-Pain), Stiffness (Lequesne-Stiffness), Physical Function (Lequesne-Physical Function), Emotional Function (Bradburn), and Social Function (MHIQ-Social). The Doyle and Lequesne indices were selected since they were developed specifically for patients with OA. The Bradburn and MHIQ indices were selected because of our familiarity with them. Finally, 3 tertiary outcome measures were used: 50' walking time, total range of movement (ROM), intermalleolar straddle. These commonly used measures of drug efficacy were selected to assess the relative efficiency of the final WOMAC battery against traditional measures, and not as supplementary measures required for validation purposes. We have not, therefore, reported statistical p values for these variables but used the data to calculate the relative efficiency of WOMAC. Individual item and aggregate item data were analyzed for each separate WOMAC dimension using both Student's t test[11] and Wilcoxon's nonparametric test[12] to assess item and dimension responsiveness (Visit 6 vs Visit 3) and the effect of parametric versus nonparametric statistical treatment of the data. Internal consistency (Visit 3) was tested using Cronbach's alpha[13], test-retest reliability (Visit 1 vs Visit 2) using Kendall's tau c statistic[14], and construct validity (Visit 3) determined using Pearson's correlation coefficient[15]. Relative efficiency was calculated (Visit 6 vs Visit 3) using the method employed by Liang, *et al:* [16] e.g., relative efficiency for WOMAC vs Walktime (WT) $= (t_{WOMAC}/t_{WT})^2$. We have not reported response data for Visit 4 (as this represented a titration step) or for Visit 5 (as this was used to assess tolerability after incremental dosing at Visit 4). However, data on these visits can be found in the paper reporting drug efficacy[4].

RESULTS

Fifty-seven patients were enrolled in the study: 28 (14 males, 14 females) received isoxicam and 29 (12 males, 17 females) received piroxicam. The mean age was 66.5 years in each group (varying from 55 to 82). The mean disease duration (i.e., symptomatology) was 8.7 years (varying from 1 to 30) in the piroxicam group and 9.3 years (varying from 2 to 26) in the isoxicam group. The knee was selected as the most severely affected joint in 39 patients (isoxicam 21, piroxicam 18) compared to the hip in 18 patients (isoxicam 7, piroxicam 11). The above differences between the 2 groups were not statistically significant. The means and standard deviations (Visit 3) for primary, secondary, and tertiary outcome measures are illustrated in Tables 2 and 3.

Pain — Responsiveness (Table 1)

Isoxicam. On Likert scaling using Wilcoxon's test, all 5 items significantly improved by Visit 6 ($p \leq 0.019$), while on VAS scaling all items achieved p values of ≤ 0.001. With the interviewer and patient global assessments and aggregate score strategies, p values of ≤ 0.001 were achieved regardless of scale (Likert vs VAS) or type of analysis (Student's t test vs Wilcoxon). When the p values derived by parametric and nonparametric analysis were compared for all 13 analyses performed using individual item = 8, aggregate

Table 1. *Summary of item content of original test form of WOMAC****

Pain†		
1	Walking	(2.58)*
2	Stair climbing	(2.62)*
3	Nocturnal	(2.63)*
4	Rest	(2.57)
5	Weight bearing	(2.51)
Stiffness†		
1	Morning stiffness	(2.52)*
2	Stiffness occurring later in the day	(2.30)
Physical Function†		
1	Descending stairs	(2.60)*
2	Ascending stairs	(2.54)*
3	Rising from sitting	(2.32)*
4	Standing	(2.64)*
5	Bending to floor	(2.51)*
6	Walking on flat	(2.40)*
7	Getting in/out car	(2.26)*
8	Going shopping	(2.40)
9	Putting on socks	(2.38)
10	Rising from bed	(2.37)
11	Taking off socks	(2.37)
12	Lying in bed	(2.36)
13	Getting in/out bath	(2.30)
14	Sitting	(2.54)
15	Getting on/off toilet	(2.67)
16	Heavy domestic duties	(2.43)
17	Light domestic duties	(2.26)
Social Function		
1	Leisure activities	(2.56)*
2	Community events	(2.15)*
3	Church attendance	(2.52)
4	With spouse	(2.65)
5	With family	(2.67)
6	With friends	(2.64)*
7	With others	(2.55)
Emotional Function		
1	Anxiety	(2.64)*
2	Irritability	(2.59)*
3	Frustration	(2.44)
4	Depression	(2.49)*
5	Relaxation	(2.39)*
6	Insomnia	(2.58)
7	Boredom	(2.62)
8	Loneliness	(2.26)
9	Stress	(2.19)
10	Wellbeing	(2.62)

* These items were duplicated on VAS scales

** These item numbers correspond to those in text and Table 4.

† Dimensions retained in final WOMAC instrument.

( ) Numbers in parentheses represent previously published[2] mean importance scores for each item. (Scale: 0=none, 1=slight, 2=moderate, 3=very, 4=extreme importance.)

Table 2. *Primary outcome measures: Visit 3 means (m) and standard deviations (s)*

Variable			Pain		Stiffness		Physical Function		Social Function		Emotional Function	
			Likert*	VAS†	Likert*	VAS†	Likert*	VAS†	Likert*	VAS†	Likert*	VAS†
WOMAC††	C	m	10.3	158.9	4.4	51.7	32.2	342.4	5.9	100.5	10.1	104.8
		s	4.4	69.0	1.8	28.5	13.8	154.1	5.5	75.8	8.5	90.1
	I	m	9.6	153.1	4.3	52.9	31.0	334.4	6.0	102.7	10.8	112.5
		s	4.0	66.3	1.6	29.7	13.5	152.6	5.2	79.5	9.6	107.3
	P	m	10.9	165.4	4.6	50.5	33.4	350.4	5.9	98.3	9.5	97.0
		s	4.7	72.6	1.9	27.7	14.3	157.9	6.0	73.3	7.5	69.3
Patient global assessment	C	m	2.5	55.3	2.4	53.3	1.9	42.1	1.1	27.2	0.7	21.0
		s	1.0	28.2	1.0	29.0	1.0	26.4	1.0	27.1	0.9	22.6
	I	m	2.3	51.7	2.3	50.1	1.8	40.1	1.1	28.5	0.8	22.8
		s	1.1	29.8	1.1	30.9	1.0	25.9	1.0	29.0	1.0	26.4
	P	m	2.6	58.6	2.4	56.1	2.0	43.8	1.1	25.9	0.6	19.4
		s	1.0	26.7	1.0	27.4	0.9	27.2	1.0	25.7	0.7	18.9
Interviewer global assessment	C	m	2.6	—**	2.1	—	2.3	—	1.1	—	1.2	—
		s	0.9	—	1.0	—	0.8	—	1.1	—	1.0	—
	I	m	2.5	—	2.2	—	2.1	—	1.1	—	1.2	—
		s	0.9	—	1.0	—	0.8	—	1.1	—	1.1	—
	P	m	2.7	—	2.0	—	2.4	—	1.0	—	1.1	—
		s	0.8	—	0.9	—	0.9	—	1.0	—	0.8	—

* Scored on values 0–4, where 0=none, 1=slight, 2=moderate, 3=very, 4=extreme.
† 0–100 mm horizontal VAS scale with terminal descriptors None and Extreme
†† Sum of WOMAC test questionnaire items: Aggregate score (pain), AI (stiffness), AI (physical), AI (social), AI (emotional).
** IgA only scored on Likert scale, *not* on VAS scale. Interviewer global assessment
C = Combined group (isoxicam + piroxicam), I = isoxicam, P = piroxicam.

Table 3. *Secondary and tertiary outcome measures: Visit 3 means (m) and standard deviations (s)*

Secondary	Combined		Isoxicam		Piroxicam	
	m	s	m	s	m	s
Bradburn total score	−3.6	3.3	−3.0	3.3	−4.2	3.2
Modified Doyle total score	2.8	1.6	2.5	1.5	3.0	1.6
Lequesne pain score	4.4	1.1	4.5	1.1	4.3	1.1
Lequesne stiffness score	1.5	0.5	1.5	0.6	1.4	0.5
Lequesne physical score	5.7	2.3	5.6	2.5	5.9	2.0
MHIQ social score	15.8	1.7	15.6	1.9	16.0	1.5
Tertiary	**Combined**		**Isoxicam**		**Piroxicam**	
	m	s	m	s	m	s
Walk time (s)	17.2	7.2	17.1	8.3	17.3	6.0
Intermalleolar straddle	79.2	18.2	80.6	17.8	77.7	18.8
ROM (°)	225.1	26.3	224.8	26.9	225.5	26.1

score = 2, interviewer global assessment = 1, and patient global assessment = 2 strategies, there was exact agreement (to 3 decimal places) in 46% of the cases, while in 54% the parametric value was smaller. The correlation coefficient between scores on Likert and VAS scales was 0.82 for patient global assessment.

Piroxicam. Regardless of the type of statistical analysis used, 80% of the items significantly improved on Likert scaling by Visit 6 ($p \leq 0.005$), but #2 failed to significantly improve. On VAS scaling using Wilcoxon's test, all 3 items achieved p values of ≤ 0.019. With respect to the interviewer and patient global assessments and aggregate score strategies, p values of ≤ 0.003 were achieved regardless of scale. When all 13 comparative analyses (individual item = 8, aggregate score = 2, interviewer global assessment = 1, patient global assessment = 2) were considered, the parametric p values were smaller in 77% of the cases and larger in 15%, while in 8% there was exact agreement. The correlation coeffi-

cient between scores on Likert and VAS scales was 0.86 for patient global assessment.

Pain — reliability. From Likert scaled responses to the 5 component items the internal consistency of the pain dimension was 0.86 for isoxicam and 0.89 for piroxicam. The corresponding values for the 3 VAS scaled responses were 0.81 and 0.73, respectively. The test-retest reliability for the combined group (i.e., isoxicam + piroxicam) was 0.68 on the Likert scale and 0.64 on the VAS scale.

Pain — validity. Higher levels of correlation (as expressed by the correlation coefficients and the proportion of items displaying a statistically significant correlation) were noted on both Likert and VAS responses between the test items and the Lequesne pain and physical function components and the Doyle Index, than between these same items and the Lequesne stiffness component, the Bradburn Index and the MHIQ social component (Table 4).

Stiffness — Responsiveness (Table 1)

Isoxicam. On Likert scaling using Wilcoxon's test, both items (morning stiffness, stiffness occurring later in the day) significantly improved by Visit 6 ($p \leq 0.004$), while on VAS scaling morning stiffness achieved a p value of ≤ 0.001 regardless of type of analysis used. With the interviewer global assessment, patient global assessment, and aggregate score strategies, p values of ≤ 0.001 were achieved regardless of scale or type of analysis. When all 7 comparative analyses (individual item = 3, aggregate score = 1, interviewer global assessment = 1, patient global assessment = 2) were considered, the p values showed exact agreement in 43% of the cases, while in 57% of the cases the parametric value was smaller. The correlation coefficient between scores on Likert and VAS scales was 0.91 for patient global assessment.

Piroxicam. On Likert scaling using Wilcoxon's test, both items significantly improved by Visit 6 ($p \leq 0.030$), while on VAS scaling, morning stiffness achieved a p value of 0.002. With respect to the interviewer and patient global assessments and aggregate score strategies, significant improvement was detected on each ($p \leq 0.013$). However, p values were smaller for patient global assessment on VAS scaling than on Likert scaling. When all 7 comparative analyses (individual item = 3, aggregate score = 1, interviewer global assessment = 1, patient global assessment = 2) were considered the parametric p value was smaller in 100% of cases. The correlation coefficient between scores on Likert and VAS scales was 0.87 for patient global assessment.

Stiffness — reliability. From Likert scaled responses to the 2 component items, the internal consistency of the stiffness dimension was 0.90 for isoxicam and 0.91 for piroxicam. Only morning stiffness was probed on the VAS scale, and, consequently, interitem reliability was not determined. Test-retest reliability for the combined group was 0.48 on the Likert scale and 0.61 on the VAS.

Stiffness — validity. The highest levels of correlation noted on both Likert and VAS scaled responses were between the 2 test items and the Doyle Index (morning stiffness $r = 0.45$, late day stiffness $r = 0.46$) and the Lequesne pain, physical function and stiffness components (morning stiffness $r = 0.22$, late day stiffness $r = 0.23$). No significant correlation was noted between the test items and the other scales (Table 4).

Physical function — Responsiveness (Table 1)

Isoxicam. On Likert scaling using Wilcoxon's test, 15 of the 17 physical function items significantly improved by Visit 6 ($p \leq 0.001$ for 4 items, $0.002 \leq p \leq 0.005$ for 8 items, and $0.008 \leq p \leq 0.009$ for 3 items). Items 10 and 14 achieved p values of 0.057 and 0.059, respectively, at Visit 6. On VAS scaling significant improvement occurred on all items, the p value being ≤ 0.001 regardless of type of analysis used. The interviewer and patient global assessments and aggregate score strategies detected significant improvements ($p < 0.003$). When p values derived by parametric and nonparametric tests were compared for all 29 analyses (individual item = 24, aggregate score = 2, interviewer global assessment = 1, patient global assessment = 2) performed, the parametric value was smaller in 62% of the cases, while in 38% of the cases there was exact agreement. The correlation coefficient between scores on Likert and VAS scales was 0.83 for patient global assessment.

Piroxicam. On Likert scaling using Wilcoxon's test, 12 of the 17 physical function items significantly improved by Visit 6 ($0.002 \leq p \leq 0.009$ for 6 items, $0.013 \leq p \leq 0.019$ for 4 items, and $0.024 \leq p \leq 0.027$ for 2 items). On VAS scaling significant improvement occurred on all items, the p value being ≤ 0.001 for 3 items, 0.006 for one item, $0.010 \leq p \leq 0.011$ for 2 items, and 0.044 for one item. With the interviewer and patient global assessments and aggregate score strategies, significant improvement was detected by Visit 6 ($p \leq 0.002$ for aggregate score and interviewer global assessment; $0.004 \leq p \leq 0.010$ for patient global assessment). When p values derived by parametric and nonparametric tests were compared for all 29 analyses performed using individual item = 24, aggregate score = 2, interviewer global assessment = 1, and patient global assessment = 2 strategies, the parametric p value was smaller in 79% of the cases, larger in 7%, while in 14% there was exact agreement. The correlation coefficient between scores on Likert and VAS scales was 0.90 for patient global assessment.

Physical function — reliability. From Likert scaled responses to the 17 component items, the internal consistency of the physical function dimension was 0.95 for isoxicam and 0.95 for piroxicam. The corresponding values for the 7 VAS scaled responses were 0.91 and 0.89, respectively. Test-retest reliability for the combined group was 0.68 on the Likert scale and 0.72 on the VAS scale.

Physical function — validity. Higher levels of correlation were noted on both Likert and VAS scaled responses between

Table 4. *Construct validity analysis: Correlation of WOMAC test items with Lequesne, Modified Doyle, Bradburn, and MHIQ indices*

Domain		Lequesne Pain	Lequesne Stiffness	Lequesne Physical	Doyle Tenderness	Bradburn Emotional	MHIQ Social
Pain							
Likert	1	(0.46/0.57	(0.14/0.35)	(0.30/0.55)	(0.25/0.46)	(−0.06/0.15)	(−0.16/−0.00)
(n=5)	2	1–5	1,4,5	1–5	1,3–5	—	—
	3	100	60	100	80	0	0
VAS	1	(0.39/0.62)	(0.04/0.24)	(0.36/0.50)	(0.36/0.57)	(−0.08/0.04)	(−0.07/0.05)
(n=3)	2	1–3	—	1–3	1–3	—	—
	3	100	0	100	100	0	0
Stiffness							
Likert	1	(0.32/0.45)	(0.22/0.23)	(0.29/0.32)	(0.45/0.46)	(−0.22/−0.09)	(−0.13/−0.08)
(n=2)	2	AMS[4], GEL[5]	—	AMS,GEL	AMS,GEL	—	—
	3	100	0	100	100	0	0
VAS	1	(0.32)	(0.27)	(0.35)	(0.47)	(−0.21)	(−0.11)
(n=1)	2	AMS	AMS	AMS	AMS	—	—
	3	100	100	100	100	0	0
Physical Function							
Likert	1	(0.15/0.51)	(−0.04/0.33)	(0.20/0.54)	(0.14/0.52)	(−0.14/0.24)	(−0.21/0.15)
(n=17)	2	1–4,6,8, 10,12–17	7,15,16	3–17	3,4,6–17	—	—
	3	77	18	88	82	0	0
VAS	1	(0.32/0.50)	(0.01/0.31)	(0.36/0.59)	(0.28/0.54)	(−0.14/0.22)	(−0.31/−0.00)
(n=7)	2	1–7	6	1–7	1–7	—	5
	3	100	14	100	100	0	14
Social Function							
Likert	1	(0.21/0.35)	(0.17/0.34)	(0.24/0.37)	(0.09/0.35)	(−0.03/0.29)	(−0.12/0.11)
(n=7)	2	1,2,5–7	2	1,2,4–7	1–3	—	—
	3	71	14	86	43	0	0
VAS	1	(0.28/0.35)	(0.22/0.37)	(0.42/0.49)	(0.36/0.46)	(−0.14/0.09)	(−0.06/0.05)
(n=3)	2	1–3	1,2	1–3	1–3	—	—
	3	100	67	100	100	0	0
Emotional Function							
Likert	1	(0.15/0.35)	(0.03/0.37)	(0.18/0.46)	(−0.04/0.20)	(0.14/0.45)	(−0.30/−0.04)
(n=10)	2	4–6,10	2,5	1–6,9,10	—	1,7–9	—
	3	40	20	80	0	40	0
VAS	1	(0.30/0.34)	(0.20/0.36)	(0.44/0.54)	(0.14/0.23)	(0.34/0.38)	(−0.25/−0.18)
(n=4)	2	1–4	4	1–4	—	1–4	—
	3	100	25	100	0	100	0

1 Min/max of Pearson correlation coefficients between individual test items and comparison indices.
2 Test item number showing statistically significant correlation with comparison indices ($p \leq 0.05$).
3 Percentage of test items showing statistically significant correlation with comparison indices.
4 AMS – stiffness after first wakening in the morning.
5 GEL – stiffness after sitting, lying, or resting later in the day.
NB – Test item numbers correspond to those identified in Table 1.

the test items and the physical component of the Lequesne Index than with these same items and the Doyle Index, the Lequesne pain and stiffness components, the Bradburn Index, and the MHIQ social component (Table 4).

Social function — Responsiveness (Table 1)

Isoxicam. On Likert scaling using Wilcoxon's test, 4 of the items achieved statistically significant improvement at Visit 6 ($p = 0.011$, 0.018, 0.033, and 0.043). On VAS scaling p values of 0.001 were achieved by the first 2 items, while a p value of 0.023 was attained by the 3rd item. Although the patient global assessment strategy resulted in a p value of 0.020 on VAS scaling, it did not statistically improve on

Likert scaling ($p = 0.110$). The aggregate score and interviewer global assessment strategies resulted in p values of ≤ 0.001 and 0.015, respectively. When all 15 comparative analyses (individual item = 10, aggregate score = 2, interviewer global assessment = 1, patient global assessment = 2) were considered, there was absolute agreement of the p value in 13% of the cases, while in 67% the parametric p values were smaller and in 20% the nonparametric p values were smaller. The correlation coefficient between scores on Likert and VAS scales was 0.87 for patient global assessment.

Piroxicam. On Likert scaling none of the items improved significantly by Visit 6 regardless of type of analysis. However, on VAS scaling item one achieved a p value of 0.002, while item 2 achieved p values of 0.010 (t test) and 0.006 (Wilcoxon). Neither the interviewer global assessment nor patient global assessment strategies detected significant improvement regardless of scale or type of analysis. However, the aggregate score strategy resulted in improvement on the VAS ($p \leq 0.008$). When all 15 comparative scale analyses (individual item = 10, aggregate score = 2, interviewer global assessment = 1, patient global assessment = 2) were considered there was exact agreement of the p value in 7% of the cases, while in 73% the parametric p values were smaller and in 20% the nonparametric p values were smaller. The correlation coefficient between scores on Likert and VAS scales was 0.62 for patient global assessment.

Social function — reliability. From Likert scaled responses to the 7 component items, the internal consistency of the social function dimension was 0.89 for isoxicam and 0.93 for piroxicam. Corresponding values for the 3 VAS scaled items were 0.89 and 0.93, respectively. Test-retest reliability for the combined groups was 0.61 on the Likert scale, and 0.59 on the VAS scale.

Social function — validity. Higher levels of correlation were noted between the test items and the Lequesne physical function and pain components (on Likert scaled responses) than with the Doyle Index or Lequesne stiffness component (Table 4). With VAS scaled responses, higher levels of correlation were noted between test items and the Lequesne physical function component, the Doyle Index, and the Lequesne pain component than with these same items and the Lequesne stiffness component. Regardless of scale, no significant correlation was noted between the test items and the Bradburn Index or with the MHIQ social component.

Emotional function — Responsiveness (Table 1)

Isoxicam. On Likert scaling using Wilcoxon's test, half of the items improved significantly by Visit 6 ($p \leq 0.043$). In contrast, all VAS scaled responses achieved p values of ≤ 0.014. Both the aggregate score and interviewer global assessment strategies showed significant improvement on Likert scaling ($p \leq 0.004$). However, the patient global assessment strategy did not detect improvement on either scale ($p \geq 0.090$). When all 19 comparative analyses (individual item = 14, aggregate score = 2, interviewer global assessment = 1, patient global assessment = 2) were considered, there was exact agreement of the p value in 10% of the cases, while in 74% the parametric p values were smaller, and in 16% the nonparametric p values were smaller. The correlation coefficient between scores on Likert and VAS scales was 0.91 for patient global assessment.

Piroxicam. On Likert scaling using Wilcoxon's test, 4 of the 10 items improved significantly by Visit 6 ($p \leq 0.050$). In contrast, the 4 VAS scaled responses all achieved p values ≤ 0.032. Although both the interviewer global assessment and aggregate score strategies demonstrated significant improvement on Likert scaling ($p = 0.004$ and $p = 0.022$, respectively), the patient global assessment strategy did not detect improvement (Likert $p = 0.779$, VAS $p = 0.187$). When all 19 comparative analyses (individual item = 14, aggregate score = 2, interviewer global assessment = 1, patient global assessment = 2) were considered, there was exact agreement of the p value in 8% of the cases while in 92% the parametric p values were smaller on Likert scaling. On VAS scaling, however, the nonparametric p values were smaller in 100% of cases. The correlation coefficient between scores on Likert and VAS scales was 0.66 for patient global assessment.

Emotional function — reliability. From Likert scaled responses to the 10 component questions, the internal consistency of the emotional function dimension was 0.96 for isoxicam and 0.91 for piroxicam. The corresponding values for the 4 VAS scaled items were 0.98 and 0.88, respectively. The test-retest reliability for the combined group was 0.72 on the Likert scale and 0.66 on the VAS scale.

Emotional function — validity. Higher levels of correlation were noted on both Likert and VAS scaled responses between the test items and the Lequesne physical function component, the Bradburn Index, and the Lequesne pain component than with the Lequesne stiffness component (Table 4). No significant correlation was noted between test items and the MHIQ social component or the Doyle Index.

Relative efficiency. When considering both treatment groups combined, 5 pain, 2 stiffness and 17 physical function items achieved statistical significance ($p \leq 0.005$) by Visit 6. Since only 3 emotional items and none of the social items achieved this level of significance, emotional and social dimensions were not subjected to relative efficiency testing (Table 5). In 83% of analyses the relative efficiency of WOMAC was > 1, i.e., more efficient than the tertiary measures. Relative efficiency values < 1 were largely accounted for by walking time scores for the piroxicam group. In 78% of comparisons the relative efficiency for VAS scaled responses was numerically greater than the corresponding Likert scaled responses.

Table 5. *Relative efficiency* of WOMAC versus tertiary outcome variables*

Tertiary Outcome Variable	Study Group	Pain		Stiffness		Physical Function		WOMAC Final Battery (FB)**	
		Likert	VAS	Likert	VAS	Likert	VAS	Likert	VAS
Walk time	C	1.4	1.5	0.7	1.4	1.2	1.3	1.4	1.5
	I	2.3	2.9	2.0	2.2	2.9	2.4	3.1	2.8
	P	0.8	0.7	0.3	0.9	0.5	0.7	0.6	0.8
Intermalleolar straddle	C	5.2	5.8	2.8	5.5	4.8	5.1	5.4	6.0
	I	2.7	3.4	2.4	2.6	3.4	2.9	3.6	3.3
	P	11.8	11.0	4.5	13.4	7.7	10.2	9.4	12.5
ROM	C	1.4	1.6	0.8	1.5	1.3	1.4	1.4	1.7
	I	1.3	1.6	1.1	1.2	1.6	1.4	1.7	1.6
	P	1.6	1.6	0.7	1.9	1.0	1.6	1.3	1.9

* Relative efficiency $= (t_1/t_2)^2$ e.g., for isoxicam $(t_{\text{pain (VAS)}}/t_{\text{walk time}})^2 = 2.9$.

** WOMAC (FB) $= AI_{\text{pain}} + AI_{\text{stiffness}} + AI_{\text{physical function}}$.

C—Combined group (isoxicam + piroxicam), I-isoxicam, P-piroxicam

DISCUSSION

In developing a new health status measure we were guided by 4 principles: adequate responsiveness, reliability and validity, and superior relative efficiency over selected traditional measures. We elected to employ a double blind, randomized, controlled parallel design since both groups of patients at Visit 3 would have a high probability of being similar with respect to their pretreatment status and response potential. Furthermore, if the 2 agents are similar in efficacy then the 2 arms of the study may be used for conducting separate tests of index responsiveness in 2 clinically equivalent groups of patients. Indeed, since no significant between-group differences were detected (Tables 2 and 3) and no significant between-drug differences were identified using the reported independent outcome measures[4], and accepting the possibility of a Type II error, nevertheless we regard the design as a legitimate and novel approach to index validation.

Responsiveness. Although isoxicam was voluntarily suspended worldwide by Warner-Lambert International in October 1985, this was not for lack of efficacy but rather for reasons of toxicity apparently related to a manufacturing problem in France[17]. Since drug efficacy was not at issue, we regard this voluntary suspension as irrelevant to the validation of WOMAC. Twenty-seven (5 pain, 2 stiffness, 17 physical, 0 social, 3 emotional) of the original 41 WOMAC items achieved statistical significance with p values ≤ 0.005 by Visit 6 for the combined group. The use of multiple analytic comparisons may result in an increase in Type I errors[18]. Even correcting for this statistical nuance, however, and accepting a high degree of covariance among Index items, the p values attained were extremely good and indicative of a high level of responsiveness for these 27 WOMAC items. Comparative analyses of nonparametric vs parametric treatment of the data suggest that while in many instances there is agreement between the 2 (and therefore that either analysis may be used), nonparametric methods may provide a more conservative estimate of the response, and for conceptual reasons relating to normality of the data, may be regarded as the preferred analytic technique. Since these observations are of limited generalizability, we are continuing to perform both parametric and nonparametric comparisons on instrument data.

Reliability. Reliability coefficients (i.e., Cronbach's alpha) of ≥ 0.80 are generally regarded as acceptable. Index items exceeded 0.85 on both VAS and Likert scales in all but one instance (pain — VAS piroxicam group = 0.73). The values achieved for test-retest reliability were somewhat lower than those for internal consistency. Nevertheless, we regard them as entirely adequate considering that (a) the test-retest interval was one week, and (b) the Kendall's tau c statistic tends to generate slightly lower coefficients of correlation[19]. We believe that the principal explanation for our lower test-retest values lies in the excessive interval (1 week) between the 2 administrations. Indeed, given the high internal consistency and sensitivity of WOMAC, and considering the constantly fluctuating symptomatology of OA, one can predict that test-retest reliability values will only be moderate. These data indicate, therefore, that all 5 WOMAC dimensions on both VAS and Likert scales are of adequate reliability.

Validity. Opinions differ as to which items should be incorporated in outcome measurement, and which numerical weights assigned to the clinical importance of different items[20]. We believe, however, that the item content of WOMAC should be generally acceptable since it is based not only on a review of both the clinimetric and OA literatures[1], but on the opinions of 100 patients with symptomatic OA who provided data on the dimensionality of their symptoms and assigned importance scores for each item subsequently used in constructing WOMAC[2]. For criterion validity testing, coefficients ≥ 0.80 are generally regarded as acceptable. However, no irrefutable gold standard cur-

rently exists against which to test criterion validity. We have, therefore, tested construct validity against other indices which probe the 5 Index dimensions of interest. Since these comparators are not gold standards, lower levels of correlation are expected. In general, however, the Index items should show a statistically significant correlation with other indices probing the same dimension (convergent construct validity). Furthermore, Index items should also show higher levels of correlation with other indices probing the same dimension than with indices probing other (particularly unrelated) dimensions (divergent construct validity). These criteria were fulfilled by the pain, stiffness and physical function components of the Index. It should be noted that since physical disability is often secondary to pain, it is not surprising that these 2 dimensions are often associated. We observed that both stiffness items showed a better correlation with the modified Doyle score than with the Lequesne stiffness component. However, we believe this to be due to the fact that the Lequesne Index probes duration of stiffness while WOMAC probes its severity. Given the interrelationship between discomfort and disability, an association between stiffness, pain, tenderness and physical function is predictable. Of note, the VAS scaled stiffness item showed a statistically significant correlation with the Lequesne stiffness component. The social component of WOMAC failed to correlate with the MHIQ social component, and although some items were reliable and responsive, this dimension was excluded from the Index. Moreover, in spite of the emotional component fulfilling construct validity criteria, and most items being reliable and responsive, we elected to withdraw the component pending a reevaluation of the social dimension. The final Index, therefore, utilizes the pain (5 items), stiffness (2 items), and physical (17 items) function subscales only (Table 1).

Relative efficiency. To be useful, a new health status measure should offer advantages over existing indices. In this respect, WOMAC offers 2 advantages. First, WOMAC and its subscales offer superior efficiency (as measured by relative efficiency scores) over selected traditional measures in assessing the efficacy of antirheumatic drugs. Such a measure, therefore, has potential for reducing sample size requirements for clinical trials using WOMAC as the primary outcome measure. Secondly, traditional measures often lack patient relevance. In contrast, WOMAC probes patient relevant outcomes, the clinical importance (Table 1) of which have been documented[2].

We believe WOMAC to be a reliable, valid, and responsive multidimensional, self-administered outcome measure designed specifically to evaluate patients with OA of the hip or knee. We are currently conducting further studies on aggregating scores across different dimensions, on the relative responsiveness of Likert and VAS scales, and the relative efficiency of WOMAC against several other health status instruments.

ACKNOWLEDGMENT

We wish to thank the following for their contribution to the development of the WOMAC: S. O'Leary, E. Grace and E. Dulai.

REFERENCES

1. Bellamy N, Buchanan WW: Outcome measurement in osteoarthritis clinical trials: The case for standardization. *Clin Rheum 1984;3:293–305.*
2. Bellamy N, Buchanan WW: A preliminary evaluation of the dimensionality and clinical importance of pain and disability in osteoarthritis of the hip and knee. *Clin Rheum 1986;5:231–41.*
3. Bellamy N, Buchanan WW: A health status instrument for osteoarthritis of hip and knee: Stage 2 analysis of a self-administered questionnaire (abstr). In: *Abstracts: Australian Rheumatism Association Annual Scientific Meeting, 1987:32.*
4. Bellamy N, Buchanan WW, Grace E: Double-blind randomized controlled trial of isoxicam vs piroxicam in elderly patients with osteoarthritis of the hip and knee. *Br J Clin Pharmacol 1986;22:149S–55S.*
5. Likert R: A technique for measurement of attitudes. *Arch Psychol 1932;140:44–60.*
6. Huskisson EC: Measurement of pain. *J Rheumatol 1982;9:768–9.*
7. Doyle DV, Dieppe PA, Scott J, *et al:* An articular index for the assessment of osteoarthritis. *Ann Rheum Dis 1981;40:75–8.*
8. Lequesne M: European guidelines for clinical trials of new antirheumatic drugs. *EULAR Bull 1980;*(suppl 6)9:171–5.
9. Bradburn NM: *The Structure of Psychological Well-Being.* Chicago: Aldine Publishing, 1969.
10. Chambers LW: *McMaster Health Index Questionnaire (MHIQ).* Hamilton: McMaster University, Department of Clinical Epidemiology and Biostatistics, 1980.
11. Colton T: Inference on means. In: *Statistics in Medicine.* Boston: Little, Brown, 1974:99–150.
12. Armitage P: Distribution-free methods. In: *Statistical Methods in Medical Research.* New York: John Wiley, 1977:394–407.
13. Cronbach LJ: Coefficient alpha and the internal structure of tests. *Psychometrika 1951;16:297–334.*
14. Conover WJ: Some methods based on ranks. In: *Practical Nonparametric Statistics.* 2nd Ed. New York: John Wiley, 1980:256–61.
15. Colton T: Regression and correlation. In: *Statistics in Medicine.* Boston: Little, Brown, 1974:189–217.
16. Liang MH, Larson MG, Cullen KE, Schwartz JA: Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. *Arthritis Rheum 1985;28:542–7.*
17. George CF: Editorial. *Br J Clin Pharmacol 1986;22:107S.*
18. Feinstein AR: Scientific decisions for data and hypotheses. In: *Clinical Epidemiology — The Architecture of Clinical Research.* Philadelphia: WB Saunders, 1985:515–7.
19. Kendall M, Stuart A: Categorized Data. In: *The Advanced Theory of Statistics,* Vol. 2. *Inference and Relationship.* 4th Ed. High Wycombe: Charles Griffin, 1979:566–615.
20. Bombardier C, Tugwell P, Sinclair A, Dok C, Anderson G, Buchanan WW: Preference for endpoint measures in clinical trials: Results of structured workshops. *J Rheumatol 1982;9:798–801.*

SECTION 4 - Exploration of Special Measurement Characteristics

Following the establishment of the basic clinimetric properties (validity, reliability and responsiveness) of the WOMAC Index, other studies and subanalyses were undertaken to explore other properties of the WOMAC Index, in particular the following: blind versus informed presentation (5,6), signal versus aggregate strategies of measurement (7,8), time frame dependency (9), and the relative responsiveness of the WOMAC Index (10,11).

Blind versus informed presentation.

In the 1980s, the measurement literature was divided as to whether patients should be shown (informed) or not shown (blind) their prior scores when self-completing a health status questionnaire. Some authorities believed that access to prior scores might bias results by creating a comparison with prior status, while others believed it provided a positive influence by allowing patients to calibrate their current status against prior scores. This issue had rarely been explored with questionnaires used in rheumatology research. One sub-analysis and one sub-study were undertaken with the WOMAC Index (5,6). The sub-analysis was part of the original pharmacologic validation study (5), while the sub-study was part of an RCT comparing flurbiprofen SR vs diclofenac sodium SR (6). No statistically significant or clinically important differences were detected in either the sub-analysis or sub-study, suggesting that changes with treatment, recorded by the WOMAC Index, were comparable using blind and informed methods of Index presentation (5,6).

Incidental to the flurbiprofen SR versus diclofenac sodium SR study, a preliminary factor analysis was undertaken of the pain and function subscales (6). Factor 1 accounted for 88% of the variance in pain and 83% of the variance in physical function. The factor loading was high on each individual pain item (0.92-0.95) and each individual physical function item (0.70-0.97). Although this was a preliminary factor analysis, based on a study of relatively small sample size, and recognizing the limitations of the methodology, the factor structure of the WOMAC Index was generally upheld.

Signal versus aggregate strategies of measurement

A recurrent theme in outcome measurement concerns whether the patient should be presented with a fixed battery of items or whether the battery should be tailored to the different symptom experience of each individual, within a predefined specification of domains or items. The potential advantages of the flexible battery are in individualisation of the questionnaire and increased index responsiveness, the potential disadvantage being in comparing "apples with oranges" in group analyses. To explore this issue with the WOMAC Index, we conducted a sub-analysis of the original orthopaedic validation study (7), and a sub-study within an RCT of tenoxicam versus diclofenac (8). Both the sub-analysis and sub-study suggested that signal measurement, where patients selected one pain, one stiffness and one physical function item from the WOMAC inventory, according to pre-specified criteria, was feasible. Both studies also indicated between-subject variability in signal selection (7,8). The signal approach was more responsive than using the entire WOMAC Index in the tenoxicam versus diclofenac study (8), but not in the orthopaedic study (7). The signal method failed to detect deterioration in non-signal items, which occurred more commonly in the tenoxicam versus diclofenac study (8) than the orthopaedic study (7). Finally, the tenoxicam versus diclofenac study (8)

suggested the occurrence of within-subject variability in signal selection over time, this observation being based on the same subjects being given a second opportunity to select a signal at the end of study. These studies suggested while signal measurement was feasible and might provide a more responsive alternative to traditional methods of measurement based on fixed item inventories, that the gain in responsiveness might be variable, and non-signal deterioration might go undetected. Collectively these data suggested that administering the entire WOMAC Index might be preferable to monitoring patient progress based only on single individualized signal items.

Time frame dependency

The most appropriate time frame over which to ask patients to recall their symptoms has rarely been studied in health status questionnaires used in rheumatology. Indeed rheumatology measures differ significantly with respect to the time period over which they require patients to rate their symptom experience. In order to explore this aspect of the WOMAC Index, a small study was conducted in which patients were presented, in random order, with three WOMAC questionnaires identical in content, but differing in their recall period (24 hours, 48 hours, two weeks) (9). No clinically important or statistically significant differences were observed between the responses to the three variations in time frame. This study provided support for varying the time frame of the WOMAC Index between one and 14 days, depending on the dynamic requirements of future studies. Ultimately, 48 hours was chosen, for conceptual reasons, as the standard time frame for the 3.1 series of WOMAC questionnaires. It was reasoned that the 48-hour time frame permitted patients more time, than the 24-hour time frame, to experience their symptoms, without becoming excessively vulnerable to recall and memory effects.

Relative responsiveness

Responsiveness or sensitivity to change is a quintessential characteristic of a health status measure for evaluating the clinical benefit of interventions in patients with OA . In addition to establishing an instrument's capacity to detect change, it is important to compare the relative responsiveness to other commonly used measures. In the two original validation studies the responsiveness of the WOMAC Index had been compared to other methods of assessment used in OA clinical trials (2,4). The comparative analyses, based on the relative efficiency statistic, suggested that the WOMAC Index was superior in responsiveness to the following observer-dependent measures: walk time, intermalleolar straddle, intercondylar distance and knee range of motion (2,4). Two additional studies were undertaken to explore the relative responsiveness of the WOMAC Index to three other standard patient self-reported health status measures (10,11). In an orthopaedic study of total knee replacement, responsiveness of the WOMAC Index was compared to that of the Health Assessment Questionnaire (HAQ) and the Arthritis Impact Measurement Scales (AIMS). The WOMAC was superior in responsiveness, as assessed by the relative efficiency statistic, in three out of four analyses (10). In a second study, conducted in an RCT of ibuprofen versus placebo, the responsiveness of the WOMAC Index was compared to the Short-Form Medical Outcomes Survey (SF-36) (11). The WOMAC was superior in responsiveness, as assessed by the effect size statistic, in 15 out of 18 analyses of pain and function scores.

Collectively these comparisons suggested that the WOMAC Index compared favourably to other commonly used measures in OA clinical research studies, and was superior in responsiveness to several of the measures.

References:

5. Bellamy, N., Goldsmith, G.H., Buchanan, W.W., Campbell, J. and Duku, E. Prior score availability: Observations using the WOMAC Osteoarthritis Index. British Journal of Rheumatology 1991;30:150-151.

6. Bellamy, N., Wells, G. and Campbell, J. Relationship between severity and clinical importance of symptoms in osteoarthritis. Clinical Rheumatology 1991;10(2):138-143.

7. Bellamy, N., Buchanan, W.W., Goldsmith, C.H., Campbell, J. and Duku, E. Signal measurement strategies: Are they feasible and do they offer any advantage in outcome measurement in osteoarthritis? Arthritis and Rheumatism 1990;33(5):739-745.

8. Barr, S., Bellamy, N., Buchanan, W.W., Chalmers, A., Ford, P.M., Kean, W.F., Kraag, G.R., Gerecz-Simon, E. and Campbell, J. A comparative study of signal versus aggregate methods of outcome measurement based on the WOMAC Osteoarthritis Index. Journal of Rheumatology 1994;21:2106-2112.

9. Griffiths, G., Bellamy, N., Kean, W.F., Campbell, J. and Gerecz-Simon, E. A study of the time frame dependency of responses to the WOMAC Osteoarthritis Index. Inflammopharmacology 1993;2:85-87.

10. Griffiths, G., Bellamy, N., Bailey, W.H., Bailey, S.I., McLaren, A.C. and Campbell, J. A comparative study of the relative efficiency of the WOMAC, AIMS and HAQ instruments in evaluating the outcome of total knee arthroplasty. Inflammopharmacology 1995;3:1-6.

11. Davies, G.M., Watson, D.J., and Bellamy, N. Comparison of the responsiveness and relative effect size of the WOMAC and SF-36 in a randomized clinical trial in patients with osteoarthritis. Arthritis Care and Research 1999;12(3):172-179.

LETTERS TO THE EDITOR

Prior Score Availability: Observations Using The WOMAC Osteoarthritis Index

SIR—Whether patients should be shown their prior scores when repeatedly self-assessing health status remains controversial [1–4]. To evaluate this issue we assessed the effect of prior score availability during a validation study of a new outcome measure for osteoarthritis (OA) [5], termed WOMAC (the Western Ontario and McMaster Universities Osteoarthritis Index). Its clinimetric properties have been reported [6–8]. We describe a further analysis comparing our experience of blind versus informed presentation of the questionnaire. We use the terms LK3.0 and VA3.0 to define certain features of the WOMAC Index.

Twenty-eight patients with OA of the hip or knee requiring NSAID entered a multicentre study comparing piroxicam with isoxicam. To be eligible, patients had to fulfil defined criteria [6]. Patients were assessed at enrolment and again 1 week later. Thereafter, patients underwent a 1-week NSAID-free washout period and were then reassessed. Finally, patients were evaluated following 2, 4 and 6 weeks of active treatment. The WOMAC questionnaire [9] was developed within a methodological framework similar to that previously discussed [10] WOMAC consists of 24 items in three dimensions: pain = 5, stiffness = 2, physical function = 17 items respectively. Subjects completed two versions of the questionnaire at each visit. The LK3.0 version required responses on 5-point Likert scales [11], and the VA3.0 version on 100 mm horizontal visual analogue (VA) scales [12]. Only 11 of the items were duplicated in the VA3.0 version (pain = 3; stiffness = 1; physical function = 7). Aggregate scores for each dimension were determined by summing the component item scores. Patients completed WOMAC blind to previous scores at all visits. At the final visit, WOMAC was initially presented to patients blind, i.e. as a blank questionnaire without access to any prior scores. Having completed WOMAC followed by three other questionnaires [6], an informed response was obtained by having patients again score their current health status on the same WOMAC questionnaire on which they had previously marked their end-of-washout scores. Descriptive statistics were calculated for disease and demographic variables. Student's *t*-test was used to compare blind versus informed scores at study termination [13]. The sign test was used to test differences in the proportion of patients underestimating or overestimating scores under blind versus informed administration of the terminal questionnaire [14]. We have made 12 statistical comparisons and, therefore, corrected the level of significance to 0.004.

Of the original 28 patients, five were withdrawn due to toxicity, concurrent illness or incomplete data. Of the remaining 23 patients, 10 were males and 13 were females; their mean age was 64.3 years, varying from 55 to 78 years. Thirteen received isoxicam and 10 piroxicam. No overall differences were noted between the two treatment groups [6], and we combined them for this analysis.

The mean values at termination for WOMAC LK3.0 are illustrated in Table I. The mode of administration resulted in different scores in 48–87% of instances but differed between dimensions. Where scores depended on the method of administration, there was a tendency to overestimate pain while underestimating stiffness and physical function under blind administration. The magnitude of such variation, expressed by subtracting blind from informed scores, is illustrated in Table II. None of these differences was statistically significant. When considering the number of subjects overestimating versus underestimating, the sign test did not detect any statistically significant difference on any dimension.

TABLE I
COMPARISON OF BLIND VERSUS INFORMED SCORES AT TERMINATION

	Termination blind	Termination informed	Difference
Pain			
LK3.0	6.56 (4.0)	6.65 (4.4)	−0.09 (2.8) −0.15, 0.881
VA3.0	85.13 (70.6)	85.13 (72.5)	0.00 (32.2) 0.00, 1.000
Stiffness			
LK3.0	3.26 (1.5)	3.04 (1.4)	0.22 (0.7) 1.55, 0.135
VA3.0	25.78 (19.3)	31.00 (25.2)	−5.22 (11.8) −2.13, 0.045
Physical function			
LK3.0	21.00 (13.7)	19.74 (12.4)	1.26 (6.4) 0.95, 0.354
VA3.0	204.39 (155.5)	197.00 (149.6)	7.39 (60.7) 0.58, 0.565

Entries are mean (SD); *t*-statistic, *P*-value.

TABLE II
NUMBER OF SUBJECTS OVER/UNDER-ESTIMATING SCORES

	Under	Same	Over	Binomial (2-tailed *P*)
Pain				
LK3.0	5	8	10	0.3018
VA3.0	9	0	14	0.4049
Stiffness				
LK3.0	8	12	3	0.2266
VA3.0	7	1	15	0.1338
Physical function				
LK3.0	12	3	8	0.5034
VA3.0	10	0	13	0.6776

Entries are number of subjects.
The sign test was used to test differences in proportions under/over-estimating scores. Since fewer than 23 differences were observed in each case, the binomial distribution was used to compute an exact significance level.

The mean values at termination for WOMAC VA3.0 are illustrated in Table I. The mode of administration resulted in different scores in 96–100% of instances but varied between dimensions. Where scores depended on the method of administration, there was a tendency to overestimate all three WOMAC dimensions under blind administration. The magnitude of such variation,

expressed by subtracting blind from informed scores, is illustrated in Table II. None of these differences was statistically significant. When considering the number of subjects overestimating versus underestimating, the sign test did not detect any statistically significant difference on any dimension.

The results of the study suggest that showing patients their previous scores made no significant difference to outcome measurement. In particular, the direction of the differences detected was not entirely predictable and varied from dimension to dimension and between the two types of scales. Furthermore, the magnitude of the difference was small, both when considering the LK3.0 and VA3.0 versions. Overall, the informed approach resulted in a more conservative estimate of the response, although the difference was neither statistically significant nor clinically important. However, this study, like its predecessors, has not established the superiority of the informed method of administration, since the appropriateness of underestimation versus overestimation was not verified by any external gold standard. Furthermore, given the dynamic nature of the study in which patients were first washed out and then treated with active therapy, we were unable to verify the claim by Guyatt and colleagues [4] that the informed mode may result in less within-patient variability and that this may decrease the sample size needed to detect changes in quality of life in clinical trials. We have, however, confirmed their observation that change with treatment is comparable using blind and informed methods of administration and have demonstrated this using a different questionnaire in a different patient population.

N. BELLAMY*†, C. H. GOLDSMITH†, W. W. BUCHANAN‡, J. CAMPBELL*, E. DUKU†

*Department of Medicine, University of Western Ontario, London, Ontario, Canada; Departments of †Clinical Epidemiology and Biostatistics, and ‡Medicine, McMaster University, Hamilton, Ontario, Canada

Received 1 November 1990

1. Jacobsen M. The use of rating scales in clinical research. *Br J Psychiat* 1965;111:545–6.
2. Hart FD, Huskisson EC. Measurement in rheumatoid arthritis. *Lancet* 1972;i:28–30.
3. Scott J, Huskisson EC. Accuracy of subjective measurements made with or without previous scores: an important source of error in serial measurements of subjective states. *Ann Rheum Dis* 1979;38:558–9.
4. Guyatt GH, Berman LB, Townsend M, Taylor DW. Should study subjects see their previous responses? *J Chron Dis* 1985;38:1003–7.
5. Bellamy N. *Osteoarthritis—an evaluative index for clinical trials.* MSc Thesis, McMaster University, Hamilton, Canada, 1982.
6. Bellamy N, Buchanan WW, Grace E. Double-blind randomized controlled trial of isoxicam vs piroxicam in elderly patients with osteoarthritis of the hip and knee. *Br J Clin Pharmacol* 1986;22:149S–55S.
7. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol* 1988; 15:1833–40.
8. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt L. Validation study of WOMAC: a health status instrument for measuring clinically-important patient-relevant outcomes following total hip or knee arthroplasty in osteoarthritis. *J Orthop Rheumatol* 1989;1:95–108.
9. Bellamy N, Buchanan WW. A preliminary evaluation of the dimensionality and clinical importance of pain and disability in osteoarthritis of the hip and knee. *Clin Rheumatol* 1986;5:231–41.
10. Kirshner B, Guyatt G. A methodological framework for assessing health indices. *J Chron Dis* 1985;38:27–36.
11. Likert R. A technique for measurement of attitudes. *Arch Psychol* 1932;140:44–60.
12. Huskisson EC. Measurement of pain. *J Rheumatol* 1982; 9:768–9.
13. Colton T. Inference on means. *Statistics in medicine.* Boston: Little, Brown and Company, 1974:134–6.
14. Armitage P, Berry G. Distribution-free methods. *Statistical methods in medical research.* 2nd ed. Oxford: Blackwell Scientific Publications, 1987:408–10.

HLA Class II Antigens in Susceptibility to Rheumatoid Arthritis

SIR—The reclassification of rheumatoid arthritis along the immunogenetic lines recently proposed by Professors Buchanan and Singal [1] is laudable in principle but flawed in practice. They suggest that HLA-DR antigens may predispose to mild RA while DQ antigens influence the severity of the disease. Furthermore, DR1 is proposed to be preferentially associated with mild RA although the evidence for this is tenuous. The results from studies conducted in Professor Singal's laboratory showed associations of mild RA with both DR1 and DR4. Those with more severe RA (i.e. requiring gold therapy) showed, not surprisingly, a still stronger association with DR4, thereby confirming the results of other studies [2, 3]. Despite this DR1 was still increased (25%) compared to random normal controls (14%) in this group with relatively severe disease despite the fact that it was already heavily weighted for DR4-bearing haplotypes (56%). The authors comment that this difference was insignificant, even if the DR4-negative patients only were considered, but this is likely to represent a type II statistical error as a result of the relatively small number of DR4-negative individuals who were available for analysis. It is noteworthy in this context that Woodrow *et al.* in their landmark paper on the immunogenetics of RA [4] calculated a very small excess risk from DR1 (1.5) which was nevertheless highly significant ($P<2.5 \times 10^{-10}$) because they analysed the combined results from many studies. The proposal that only those subtypes of DR4 capable of presenting a particular MHC/ligand epitope (also shared with DR1 and possibly DRw10) are associated with RA has now been rigorously tested [5]: the shared epitope may be found in at least 85% of patients. The relatively smaller association of DR1 with severe RA than mild RA claimed by Buchanan and Singal [1] is a direct consequence of the preferential association of DR4 with severe disease [2, 3]. Since the fall-off in DR1 with increasing disease severity is therefore a secondary phenomenon it clearly does not constitute a firm basis for any diagnostic recategorization.

The role of DQ in RA has been incompletely evaluated so far. Certainly the extraordinarily strong association of DQw7 reported by Singal *et al.* [6] in moderately severe RA has not been confirmed by other studies in patients requiring disease-modifying drugs [5]. The most recent data from Professor Singal's group [1] are statistically flawed since the frequency of DQw7 in the severe RA group (93%) is contrasted with that in a random control group (19%) despite the fact that the former is highly selected for the DQw7- and DQw8-bearing DR4 haplotypes. Any increase in DQw7 could also be accounted for by linkage disequilibrium with DR4:Dw4 (see below). Our results of DQ DNA haplotyping previously reported [5] have now been extended to a larger group of 119

138-143 *Clinical rheumatology*, 1991, 10, N° 2

Relationship between Severity and Clinical Importance of Symptoms in Osteoarthritis

N. BELLAMY, G. WELLS*, J. CAMPBELL

Summary **Seventeen patients with primary osteoarthritis of the knee were evaluated with respect to the severity and clinical importance of pain, stiffness and physical function during the conduct of a double-blind randomized controlled trial of flurbiprofen SR versus diclofenac sodium SR using the WOMAC Osteoarthritis Index. Mean importance scores were similar for items within the same dimension as well as between items in different dimensions. In general, low levels of correlation were noted between the severity and importance of symptoms. Analysis of individual WOMAC items within a given subscale indicated that, although highly correlated, they differed from one another. Factor analysis further supported the contention that scores from items within a subscale could be summated into subscale scores. These observations are of importance in the weighting and aggregation of items within discrete dimensions and have the potential for reducing sample size requirements for clinical trials in osteoarthritis.**

Key words: **Clinical Metrology, Importance, Osteoarthritis.**

INTRODUCTION

The principal objective of outcome measurement procedures for therapeutic trials of nonsteroidal anti-inflammatory drugs (NSAIDs) is to detect statistically significant, clinically important, differences in health status between competing treatment programmes. Although much attention has been focused on outcome measures for rheumatoid arthritis (RA) (1,2), much less attention has been paid to the study of patients with osteoarthritis (OA). In general, clinical investigators and international agencies have recommended the use of multiple outcome measures for OA trials (3-6). The use of multiple outcome measures, however, necessitates a downward correction in the statistical p value which results in increased sample size requirements (7). Such problems can be overcome by weighting and aggregating different measures into a single composite index (8). Such a procedure, however, requires respect to relative clinical importance of different items, as well as differences in the lengths of the scales on which the different components are measured. Smythe et al have constructed a composite index, termed the Pooled Index, for application in patients with RA (9). However, their statistical techniques, while correcting for variability in scale length, do not respect the relative clinical importance of the component items (9). In contrast, Gade (10) and Freeman et al (11) have suggested diametrically opposed weighting and aggregation systems for assessing range of movement. We have recently conducted a series of studies (12-16) validating a tridimensional self-administered questionnaire probing pain, stiffness and physical function in patients with OA of the hip or knee. The resulting index is termed the Western Ontario and McMaster Universities (WOMAC) OA Index. An earlier study (13) had indicated that the twenty-four component questions of the Index were regarded by symptomatic patients as being of similar mean clinical importance. In that study (13) importance was measured on 5-point Likert (17) scales. We were unable to determine the extent to which the severity of the patients' symptoms influenced their determination of importance scores. Since we are now using a battery of 10 cm visual analogue (VA) (18) scales as the scaling base for the WOMAC Index, and since we wish to aggregate the component questions within, and, if possible, across the three subscales, we have administered the WOMAC instrument during the conduct of a double-blind randomized controlled multi-centre trial of flurbiprofen (Ansaid-SR) versus diclofenac (Voltaren SR) in OA knee. The study had two major objectives: 1) To examine the relationship between importance and severity scores using WOMAC; 2) To examine whether

Department of Rheumatology, University of Western Ontario, London, Ontario, Canada; *Division of Biometrics, Health and Welfare Canada, Ottawa, Ontario, Canada.

Table I: *Pain: Mean severity score (standard deviation) with blind and informed administration; mean importance score (standard deviation); Pearson correlation coefficient and p-value of severity and importance with blind administration*

Item	Score-Blind		Score-Informed		p-value	Correlation	Importance		Correlation-Blind		p-value t-test from average
	Mean	s.d.	Mean	s.d.	paired t-test	Blind/Informed	Mean	s.d.	r	p	
Walking on a flat surface	41.35	27.21	44.24	30.18	.37	.91	77.88	15.24	.23	.37	.54
Going up or down stairs	52.18	26.67	54.59	28.19	.43	.90	81.35	11.78	.05	.85	.00
At night while in bed	37.18	27.86	38.94	30.00	.34	.97	71.18	21.76	.31	.23	.02
Sitting or lying	37.53	29.84	40.94	28.96	.14	.95	71.18	22.92	.17	.52	.05
Standing upright	45.06	30.21	47.12	28.14	.35	.96	67.77	26.77	.60	.01	.38

a simple addition of component item scores into three separate subscale scores was justified. In order to ensure generalizability of our observations to different modes of index presentation, we compared severity scores at termination with prior scores both unavailable (blind presentation) and available (informed administration).

MATERIALS AND METHODS

Seventeen patients attending the rheumatology outpatient clinic at Victoria Hospital, London, with definite radiographic and clinical evidence of primary OA knee were entered as part of a multi-centre double-blind randomized controlled trial comparing Ansaid-SR with Voltaren SR. To be eligible patients had to fulfil the following criteria: Inclusion criteria: symptoms requiring NSAID medication, age ⩾18 years, symptoms ⩾2 months, informed consent obtained; Exclusion criteria: Gastrointestinal ulceration or bleeding, NSAID hypersensitivity, significant uncontrolled impairment of major organ function, pregnancy or lactation, concomitant use of lithium or anticoagulants, clinically significant abnormalities in haematology or biochemistry. Following enrollment, patients underwent a 3-7 day washout period during which only acetaminophen was allowed. Subsequently, patients were randomly allocated to receive either Ansaid-SR (200 mg po once daily) or Voltaren SR (100 mg po once daily) for six weeks. The medications were identical in appearance thus maintaining physician and patient blinding. Patients were assessed at the end of Week 3 and Week 6. In addition to the WOMAC instrument, data were collected on several other variables. It should be noted, however, the WOMAC Index was only applied in our centre and that this report is confined to severity versus importance issues of the WOMAC instrument in our 17 patients. Data collected from other locations in this multi-centre study, as well as comparison of the two drugs for efficacy and tolerability, will be reported in a separate publication by the other investigators. At the end of the study patients completed all three subscales of the WOMAC Index rating the severity of their symptoms on 10 cm horizontal VA scales (terminal descriptors: None, Extreme) first without their prior WOMAC scores being available (i.e. blind), and again, some five minutes later with their prior scores available (i.e. informed) (13). The reliability, face, content validity, construct validity, and responsiveness of each of the 24 questions posed have been previously defined, verified and reported (14,15). After another five minutes, patients were shown an alternate form of WOMAC, in which they were asked to separately rate on 10 cm horizontal VA scales (terminal descriptors: None, Extreme) the importance, which they attached to being completely symptom free of each of those 24 symptoms. From these data, the mean and standard deviation for severity and importance scores of each item at study termination were calculated. For severity scores, these parameters were calculated for both blind and informed assessments. To examine the relationship between severity and importance, Pearson correlation matrices were constructed for each individual item and the level of correlation and statistical significance determined. The effect of administering the WOMAC Index under blind and informed conditions (i.e., prior score availability) was examined by comparing the mean severity scores under both types of administration using Student's t-test. The issue of whether a simple addition of component items to form subscale totals was examined using Student's t-test, correlation coefficients and factor analysis techniques. Since the 24 component items of WOMAC were considered in each of the three different statistical analyses, the p value, defining statistical significance, was corrected downward by a factor of 24 resulting in a value of $<.002$ (7).

RESULTS

The results are summarized in Tables I-III. The mean age of the study population was 60.24 years (range = 52 - 65) and the mean disease duration 8.57 years (range = 8 months - 20 years). There were 7 male and 10 female subjects. Their radiographic ratings according to

Table II: *Stiffness: Mean severity score (standard deviation) with blind and informed administration; mean importance score (standard deviation); Pearson correlation coefficient and p-value of severity and importance with blind administration.*

Item	Score-Blind		Score-Informed		p-value	Correlation	Importance		Correlation-Blind		p-value t-test from average
	Mean	s.d.	Mean	s.d.	paired t-test	Blind/Informed	Mean	s.d.	r	p	
Morning	42.47	30.01	42.65	33.05	.94	.96	66.71	25.01	.41	.11	.74
Gelling	43.88	28.62	43.94	30.70	.98	.93	66.88	22.20	.43	.09	.74

the Atlas of Standard Radiographs (19) were as follows: Grade I = 3, Grade II = 4, Grade III = 6, Grade IV = 4. The functional status ratings according to the Steinbrocker classification (20) were as follows: Grade II = 13, Grade III = 4. Of the seventeen patients, 8 received Ansaid-SR and 9 received Voltaren SR. The range of possible values for severity and importance on the VA scales was 0-100 mm.

Blind versus informed administration

For *blind* administration the range of severity scores calculated from the component questions of each dimension was as follows: Pain = 37.18 - 52.18 (Table I); Stiffness = 42.47 - 43.88 (Table II); Physical Function = 34.24 - 60.82 (Table III). For *informed* administration, the range of severity scores was as follows: pain = 38.94 - 54.59 (Table I); Stiffness = 42.65 - 43.94 (Table II); Physical Function = 35.77 - 63.65 (Table III). No statistically significant differences were noted between the severity scores at termination under blind versus informed administration (Tables I - III). The item scores for the two forms of administration were highly correlated (Tables I - III). Since there was no difference between severity scores obtained by blind and informed administration, we have reported the importance issue only with respect to blind administration.

Importance scores

Mean importance scores for component items were as follows: Pain = 67.77-81.35 (Table I), Stiffness = 66.71-67.88 (Table II), Physical Function = 56.29-76.24 (Table III). In all but one instance (bending to floor), the standard deviation for importance scores was less than for the corresponding severity scores.

Severity versus importance scores

We examined the relationship between severity scores and importance scores using correlation coefficients.

Table III: *Physical Function: Mean severity score (standard deviation) with blind and informed administration; mean importance score (standard deviation); Pearson correlation coefficient and p-value of severity and importance with blind administration*

Item	Score-Blind		Score-Informed		p-value	Correlation	Importance		Correlation-Blind		p-value t-test from average
	Mean	s.d.	Mean	s.d.	paired t-test	Blind/Informed	Mean	s.d.	r	p	
Descending stairs	49.53	28.56	49.29	28.70	.96	.83	71.77	18.36	.30	.24	.19
Ascending stairs	51.29	25.39	54.41	25.64	.29	.89	71.24	17.53	.20	.43	.18
Rising from sitting	46.77	26.94	50.06	31.16	.55	.71	66.29	24.93	.22	.39	.24
Standing	43.18	28.31	46.29	28.34	.24	.93	68.82	27.35	.53	.03	.47
Bending to floor	48.24	30.22	48.41	32.22	.94	.95	59.65	30.75	.64	.01	.25
Walking on flat surface	44.94	28.18	46.35	28.14	.49	.93	76.24	21.93	.34	.20	.63
Getting in/out of car	49.12	26.76	48.94	27.15	.94	.93	72.06	18.38	.36	.15	.16
Going shopping	48.77	30.33	53.35	29.24	.04	.96	68.77	25.84	.15	.56	.12
Putting on socks	42.77	31.44	42.88	32.68	.94	.98	61.41	30.06	.54	.03	.69
Rising from bed	36.53	31.18	40.65	32.31	.03	.98	56.29	29.10	.62	.01	.01
Taking off socks	41.41	31.97	41.47	31.35	.98	.96	58.53	27.69	.49	.05	.49
Lying in bed	34.24	28.31	38.53	29.69	.07	.95	60.35	20.89	.59	.01	.01
Getting in/out of bath	53.12	30.90	53.82	31.60	.71	.97	71.65	21.57	.47	.06	.09
Sitting	35.59	29.32	35.77	29.08	.94	.94	59.24	23.51	.69	.00	.00
Getting on/off toilet	37.65	29.89	41.35	30.37	.08	.97	63.65	28.01	.39	.13	.03
Heavy domestic duties	60.82	27.91	63.65	28.72	.30	.93	75.47	21.66	.48	.05	.00
Light domestic duties	38.88	29.61	43.35	31.72	.11	.94	71.59	27.56	.23	.38	.02

The following guidelines were used to interpret correlation coefficients: poor correlation = 0 < 0.3; moderate correlation 0.3 < 0.6; good correlation 0.6 < 0.8; excellent correlation ≥0.8. Seven coefficients were poor, 13 moderate, 4 good but none were excellent. No statistically significant correlation was noted between the importance and severity scores for any of the 24 WOMAC items.

Item aggregation

Using Student's t-test no significant difference was detected between the scores of individual items and the average score for the subscale to which that item belonged, except in two instances of physical function (sitting, heavy domestic duties) (Tables I - III). The level of interitem correlation for components of each of the three subscales was high: pain = 0.79-0.96, stiffness = .83, physical function = 0.52-0.98. Most correlation coefficients were ≤0.80. Principal component analysis was not performed for stiffness because the subscale contains only two items. However, analysis of the pain and physical function subscales showed that Factor I accounted for 88% of the variance in pain and 83% of the variance in physical function. The factor loading was high on each individual pain item (0.92-0.95) and each individual physical function item (0.70-0.97). There was relatively little additional variance accounted for by Factor II (pain = 7%, physical function = 6%).

DISCUSSION

In this study we have defined the severity and importance of 24 different symptoms of knee OA using the WOMAC Osteoarthritis Index. There is controversy in the literature as to whether serial questionnaires should be administered with or without access to prior scores (21,22). The conservative view prefers blind administration. Since we detected no difference between severity scores obtained by blind versus informed administration of the index in this study, we have based our report on blind administration. However, we have performed parallel analyses using informed scores, obtaining similar results and no interpretative differences.

In a previous study using a Likert-scaled version of WOMAC, we noted that the importance ascribed to symptoms was similar for different items in the same dimensions, as well as for symptoms in different dimensions. If direct comparison can be drawn between Likert and VA scaled responses, then it is of note that the mean importance scores on VA scales, in this study, 56.29 - 81.35 (i.e., 56%-81% along the length of the scale), were similar to mean importance scores reported 2.26 - 2.69 (i.e. 57%-66% along the length of the scale) on the 5-point Likert scales (0 = none, 1 = slight, 2 = moderate, 3 = very important, 4 = extremely important) for the 24 items in our previous study (13). Thus, given that 2 is the mid-point of the Likert scale, and 50 the mid point of the VA scale, we interpret our data as indicating that the majority of patients rate their symptoms somewhere between moderate and very important, and that there is a relatively narrow range for such values. These data support the contention that symptomatic patients regard their own particular symptoms of similar importance to those of other patients.

We had originally considered the possibility of using differences in importance scores as a method of weighting subscale items in the WOMAC Index. From the correlation analysis of severity versus importance, it can be seen that these two elements are distinct and require separate consideration. From a conceptual standpoint, the similarity in importance scores would suggest that items could be simply added together. We wish, however, to explore the statistical justification for such a system of weighting and aggregation. The fact that several items differed significantly from the subscale average, suggests that the items measure different aspects of the dimension and that all were relevant in aggregation. Likewise, although the factor analysis was only conducted on 17 subjects, the high percent of variability accounted for by Factor I and the very high Factor loading on each individual component item, further supports the contention that there are no redundant items in the WOMAC inventory. The high interitem correlation noted within each subscale and the fact that every single item had a high factor loading support the assumption that WOMAC subscale scores for pain, stiffness and physical function can be derived by the simple process of addition. The practical applications of our observations are as follows: 1) The fact that individual patients ascribe moderate levels of importance to each of the 24 WOMAC symptoms provides adequate justification for routinely measuring these symptoms as outcomes in clinical trials provided they fall within the dimensionality of the potential response to the intervention (i.e. NSAID therapy). Other investigators have suggested that the measurement process should focus on clinically relevant outcomes (23), and, indeed in this, as well as our previous study (13), we have demonstrated the clinical relevance of the WOMAC question inventory; 2) The different methods of analysis employed suggest that each item carries the same weight and that the three subscale scores can be derived by the process of simple addition; 3) If some dimensions carry consistently greater importance, then they should also carry more weight in the

construction of any composite index; 4) Aggregation has important implications for sample size requirements for clinical trials. For example, without aggregation 24 statistical tests of independent variables would necessitate a reduction in the Type I error from .05 by 24 fold (i.e. $p < .002$). However, aggregation of the WOMAC inventory within each of its three dimensions would necessitate as maximum Type I error correction from .05 by only a factor of 3 (i.e. $p \leq .017$). Furthermore, the construction of a composite index, which combined the pain, stiffness and physical function subscales into a single value, would result in only a single statistical comparison and obviate the need for any Type I error correction below .05. The standard formula for calculating sample size for clinical trials is as follows: n per group = $2\left[\frac{(Z_\alpha + Z_\beta)\sigma}{\Delta}\right]^2$, where σ = standard deviation, and Δ = the change the investigator is interested in detecting (24). As the value for the Type I error is reduced, the value of Z_α increases and the sample size requirements for a proposed trial increase correspondingly.

In this study we have demonstrated that there is no difference in termination scores between blind and informed methods of administration of the WOMAC Osteoarthritis Index. Although symptomatic patients regard their symptoms of similar importance regardless of severity, our observations suggest that importance and severity are little associated. We have also shown that there are no redundant items in the WOMAC Index and demonstrated a justification for deriving subscale scores by the simple addition of component items. We did not address the issue of whether the three separate dimensions can be aggregated into a single total score in this study. This issue is the subject of a current study. At present we recommend that WOMAC subscale scores be constructed by the simple aggregation of items within each of the three different dimensions and that any comparative analysis treats each dimension as a separate entity. For the definitive studies we recommend setting the Type I error at $\leq .017$ to make adequate correction for multiple comparisons. When the instrument is being used for pilot studies, however, we do not recommend any correction and prefer to set the Type I error at ≤ 0.05. We make this differentiation to respect the scientific rigour of a definitive study and to reflect the view of Dr. A. Feinstein that the purpose of a fishing expedition is to catch fish.

Acknowledgments: This work was supported in part by the Arthritis Society of Canada, and in part by a grant from the Upjohn Company of Canada.

REFERENCES

1. Bombardier, C., Tugwell, P., Sinclair, A., Dok, C., Anderson, G., Buchanan, W.W. Preference for endpoint measures in clinical trials: results of structured workshops. J Rheumatol 1982, 9, 5, 798-801.
2. Spector, T.D. Epidemiological aspects of studying outcome in rheumatoid arthritis. Br J Rheumatol 1988, 27 (Suppl. I), 5-11.
3. Altman, R.D., Meenan, R.F., Hochberg, M.C., Bole, G.G.Jr., Brandt, K., Cooke, T.D.V., Greenwald, R.A., Howell, D.S., Kaplan, D., Koopman, W.J., Mankin, H., Mikkelsen, W.M., Moskowitz, R., Sokoloff, L. An approach to developing criteria for the clinical diagnosis and classification of osteoarthritis: A status report of the American Rheumatism Association Diagnostic Subcommittee on Osteoarthritis. J Rheumatol 1983, 10, 2, 180-183.
4. Bellamy, N., Buchanan, W.W. Outcome measurement in osteoarthritis clinical trials: the case for standardisation. Clin Rheumatol 1984, 3, 3, 293-303.
5. Guidelines for the clinical evaluation of anti-inflammatory and antirheumatic drugs (adults and children). US Department of Health and Human Services, Public Health Service, Food and Drug Administration, April 1988, 12-15.
6. Guidelines for the clinical investigation of drugs used in rheumatic diseases. World Health Organization Regional Office for Europe, European League Against Rheumatism, 1983, 1-33.
7. Feinstein, A.R. Scientific decisions for data and hypotheses. In: Clinical Epidemiology - The Architecture of Clinical Research. Philadelphia, W.B. Saunders Company, 1985, 515-517.
8. Stewart, A.L., Ware, J.E.Jr., Brook, R.H. Advances in the measurement of functional status: Construction of aggregate indexes. Med Care 1981, XIX(5), 473-488.
9. Symthe, H.A., Helewa, A., Goldsmith, C.H. "Independent assessor" and "Pooled Index" as techniques for measuring treatment effects in rheumatoid arthritis. J Rheumatol 1977, 4, 2, 144-152.
10. Gade, H.G. A contribution to the surgical treatment of osteoarthritis of the hip joint: A clinical study. III. Comments on the follow-up examinations and the evaluation of the therapeutic results. Acta Chir Scand (Suppl) 1947, 120, 37-45.
11. Freeman, M.A.R., Swanson, S.A.V., Todd, R.C. Total replacement of the knee using the Freeman-Swanson knee prothesis. Orthop Related Res 1973, 94, 153-170.
12. Bellamy, N. Osteoarthritis-An Evaluative Index for Clinical Trials. M. Sc. Thesis, McMaster University, Hamilton, Canada, 1982.
13. Bellamy, N., Buchanan, W.W. A preliminary evaluation of the dimensionality and clinical importance of pain and disability in osteoarthritis of the hip and knee. Clin Rheumatol 1986, 5, 2, 231-241.
14. Bellamy, N., Buchanan, W.W., Goldsmith, C.H., Campbell, J., Stitt, L. Validation study of WOMAC: A health status instrument for measuring clinically-important patient-relevant outcomes following total hip or knee arthroplasty in osteoarthritis. J Orthop Rheumatol 1988, 1, 95-108.
15. Bellamy, N., Buchanan, W.W., Goldsmith, C.H., Campbell, J., Stitt, L.W. Validation study of WOMAC: A health status instru-

ment for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988, 15, 12, 1833-1840.
16. Bellamy, N., Goldsmith, C.H., Buchanan, W.W., Campbell, J., Duku, E. Prior score availability: Observations using the WOMAC Osteoarthritis Index. Br J Rheumatol 1991, 30, 150-151.
17. Likert, R., A technic for the measurement of arthritis. Arch Psychol 1932, 140, 44-60.
18. Scott, J., Huskisson, E.C. Graphic representation of pain. Pain, 1976, 2, 175-184.
19. Kellgren, J.H., Lawrence, J.S. Radiological assessment of osteoarthrosis. Ann Rheum Dis 1957, 16, 494-502.
20. The Cooperating Clinics Committee of the American Rheumatism Association. A seven-day variability study of 499 patients with peripheral rheumatoid arthritis. Arthritis Rheum 1965, 8, 2, 302-334.
21. Guyatt, G.H., Berman, L.B., Townsend, M., Taylor, D.W. Should study subjects see their previous responses? J Chron Dis 1985, 38, 12, 1003-1007.
22. Scott, J., Huskisson, E.C. Accuracy of subjective measurements made with or without previous scores: an important source of error in serial measurement of subjective states. Ann Rheum Dis 1979, 38, 558-559.
23. Tugwell, P., Bombardier, C., Buchanan, W.W., Goldsmith, C.H., Grace, E., Hanna, B. The MACTAR patient preference disability questionnaire - an individualized functional priority approach for assessing improvement in physical disability in clinical trials in rheumatoid arthritis. J Rheumatol 1987, 14, 3, 446-451.
24. Colton, T. Inference on Means. Statistics in Medicine, Boston. Little, Brown and Company, 1974, p. 142.

Received: 5 May 1989; Revision-accepted: 12 November 1990
Correspondence to: Dr. N. BELLAMY,
Suite 402A, Westminster Tower, Victoria Hospital, 800 Commissioners Road East, London, Ontario, Canada N6A 4G5.

BRIEF REPORT

SIGNAL MEASUREMENT STRATEGIES: ARE THEY FEASIBLE AND DO THEY OFFER ANY ADVANTAGE IN OUTCOME MEASUREMENT IN OSTEOARTHRITIS?

NICHOLAS BELLAMY, W. WATSON BUCHANAN, CHARLES H. GOLDSMITH, JANE CAMPBELL, and ERIC DUKU

The applicability of a signal measurement strategy was compared with a traditional method of measuring outcome in osteoarthritis. The signal method detected statistically significant alterations in health status with small sample sizes and with a relative efficiency close to or at unity. The prevalence of deterioration in nonsignal items was low. Signal methods of measurement may provide an alternative approach to outcome measurement in osteoarthritis clinical trials.

Signal measurement is a methodologic technique in which the measurement of disease is based on well-defined, individualized targets. Thus, in arthritis, measurement is restricted to only 1, or a few, selected joint(s) or symptom(s). This technique has 2 objectives: 1) to tailor the measurement process to the symptom profile of the individual patient, and 2) to improve the efficiency of the measurement process by excluding joints or other items that lack response potential. In spite of the possible advantages of signal measurement and the success of the technique as reported by Dixon et al (1), it is noteworthy that neither Ward et al (2) nor Egger et al (3) demonstrated any clear superiority of this technique over more traditional methods. However, Tugwell et al (4) recently replicated the success of a signal technique in potentially reducing sample size requirements for clinical trials.

Since all of the above-mentioned studies were conducted with rheumatoid arthritis patients, the generalizability of the observations to the measurement of disease in osteoarthritis (OA) patients is unknown. For this reason, during the validation of a new outcome measure for OA clinical trials, designated the Western Ontario and McMaster Universities (WOMAC) Osteoarthritis Index (5), we compared the performance of a signal method of measurement with that of a measure based on aggregated items, in each of 3 different dimensions. The reliability, validity, and responsiveness of the WOMAC Index have been documented (5).

This report presents a further analysis of data from that study, comparing signal measurement versus aggregate techniques of measurement. Specifically, we wished to examine whether there was any advantage in using the WOMAC question inventory as a menu for identifying signal symptoms for each individual in each of 3 dimensions, compared with administering the inventory in its entirety. To address this, we identified (a) the nature and severity of symptoms in the WOMAC inventory that were selected as signals, (b) whether the signal strategy could be successfully applied in outcome measurement in OA of the hip or knee, (c) the relative efficiency of the 2 techniques, (d)

From the Department of Medicine, University of Western Ontario, London, Ontario, and the Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada.

Supported in part by a grant from the Arthritis Society of Canada and a grant from Parke-Davis Canada Inc.

Nicholas Bellamy, MD, MSc, FRCP (Edin, Glas, and C), FACP: Associate Professor of Medicine, Epidemiology, and Biostatistics, University of Western Ontario; W. Watson Buchanan, MD, FRCP (Edin, Glas, and C): Professor of Medicine, McMaster University; Charles H. Goldsmith, PhD: Professor of Clinical Epidemiology and Biostatistics, McMaster University; Jane Campbell, BA: Research Assistant, University of Western Ontario; Eric Duku, BSc, Grad Dip, MSc: Department of Clinical Epidemiology and Biostatistics, McMaster University.

Address reprint requests to Nicholas Bellamy, MD, MSc, Suite 402A, Victoria Hospital, Westminster Tower, 800 Commissioners Road East, London, Ontario N6A 4G5, Canada.

Submitted for publication August 22, 1989; accepted in revised form December 5, 1989.

whether deterioration in nonsignal items was overlooked using the signal strategy, and (e) the sample size implications of the 2 techniques.

PATIENTS AND METHODS

Thirty patients with primary OA of the hip or knee were interviewed prior to total joint replacement surgery and at 6 weeks, 3 months, and 6 months after surgery. The criteria used for patient selection, as well as the disease and demographic profiles of the study group, have been reported previously (5).

At the initial interview, each patient was shown the item inventory of the WOMAC Index and asked to select 1 item from each of the 3 dimensions that was of importance to him or her as the focus for further measurement during the study. Specifically, patients were asked to select 1 pain item, 1 stiffness item, and 1 physical function item that was most important to them, i.e., that they most hoped the treatment they were about to receive would improve. However, in order to compare signal and aggregate strategies at each of the assessment points, the full WOMAC Index was self-administered throughout the study, thus obtaining serial measurements on each of the components of its 3 dimensions.

The present analysis was confined to a comparison of baseline results versus results at 6 months, because these represent the extremes of the response for both signal and aggregate strategies. Since we wished to compare the scores of the same individuals at these 2 time points using both parametric and nonparametric statistics for paired data, we restricted our analysis to those patients who both completed the 6-month study and completed all WOMAC questionnaires in their entirety. This reduced the available sample for analysis to 20 patients for the pain dimension, 27 for the stiffness dimension, and 24 for the physical function dimension. The remaining questionnaires could not be analyzed using paired statistics because they were not entirely complete.

All responses in this study were made on 5-point Likert scales (0 = none, 1 = mild, 2 = moderate, 3 = severe, and 4 = extreme). For pain and stiffness, these scales required responses to questions about severity, while for physical function, the questions were phrased in terms of degree of difficulty experienced. Aggregate values for each dimension were calculated by totaling the scores across all component items. Signal values for each dimension were represented by the score of the individual item selected.

To examine the possibility that patients tended to focus on signals pertaining to the aspects of their physical condition that were most severely affected, the relative ranking of the signal measurement among other items in the same dimension was determined. Values were tied for some items; that is, signal values had the same severity score as nonsignal items. Such occurrences are indicated in the tables. In instances where signal severity scores were lower than scores for some nonsignal items, the rank is specified, regardless of any ties in higher-ranked items (for ranking rules, see Table 1).

Table 1. Pain dimension signals*

	No. of patients selecting item as a signal	Item ranking†	Signal score at baseline		Score at baseline		Change from baseline to 6 months (*P*)	
			Mean‡	SD	Mean‡§	SD	*t*-test (parametric)	Wilcoxon test (nonparametric)
Individual item								
Pain while walking on a flat surface	5	3, 1¶, 1¶, 1¶, 1¶	2.8	1.3	2.1	1.2	<0.001	0.001
Pain while walking up or down stairs	10	1¶, 1¶, 1¶, 1, 1¶ 2, 1¶, 1, 1, 1	2.8	0.8	2.6	0.9	<0.001	<0.001
Pain at night while in bed	3	1¶, 1¶, 1¶	3.3	0.6	1.9	1.4	<0.001	0.001
Pain while sitting or lying	1	3¶	2.0	ND	1.6	1.4	0.002	0.004
Pain while standing upright	1	1¶	4.0	ND	2.4	1.0	<0.001	<0.001
Signal (n = 20)	NA	NA	2.9	0.9	2.9	0.9	<0.001	<0.001
Aggregate (n = 20)	NA	NA	10.5	4.4	10.5	4.4	<0.001	<0.001

* ND = not determined (due to lack of sufficient numbers); NA = not applicable.
† Item rank for each patient who selected item as a signal (possible number of ranks = 5).
‡ Except for the aggregate, possible scores could range from 0 (not affected) to 4 (most severe).
§ Derived by aggregating all scores at baseline for each individual item.
¶ Tied values. Note: In the case of a tied pair, there is a discontinuity of 1 category in the subsequent ranking. Thus, if there is a tied pair third rank, then the 5 ranks are expressed as follows: 1, 2, 3¶, 5.

Each of the 3 dimensions of WOMAC was analyzed separately using the Wilcoxon matched pairs signed rank test and Student's paired *t*-test (6). The SPSS-X software program was used to calculate the test statistics (7). *P* values less than 0.05 were considered significant, and no correction was made for multiple comparisons. We elected to use both parametric and nonparametric techniques, since parametric techniques may be applicable for certain ordinal-level data. However, our data were generally not normally distributed, and we believe the use of the nonparametric technique provides a more conservative estimate of statistical significance (6); results of 2 previous studies using the WOMAC Index support this contention (5,8).

Relative efficiency (RE) for the parametric analysis was calculated using the method reported by Liang et al (9), where RE (signal versus aggregate) = $(t_{signal}/t_{aggregate})^2$. For the nonparametric tests, RE (signal versus aggregate) = $(Z_{signal}/Z_{aggregate})^2$. Nonsignal deterioration was determined by comparing presurgery and postsurgery scores for each item not identified as a signal in the full WOMAC inventory and by noting the frequency and magnitude of any deterioration. Sample size requirements, based on matched pairs analysis, were calculated for both the signal and aggregate strategies for each dimension, using a parametric technique. Calculations were based on the assumptions that *P* values less than 0.05 were significant, the power of the test was 90%, and the difference to be detected could be in either a positive or a negative direction, i.e., a 2-tailed test of the null hypothesis. For each strategy, the sample size formula used was as follows: n matched pairs of observations = $([Z_{0.05} + Z_{0.10}]\sigma/D)^2$, where σ = the standard deviation of differences and D = the decimal difference (from baseline) to be detected (i.e., 0.25 of mean) (6).

RESULTS

Pain. Each of the 5 items included in the pain dimension was selected as a signal by 1 or more patients (Table 1). Measures of pain under conditions of activity were more frequently selected as signals than those under conditions of passivity: Pain observed when walking up or down stairs was the most frequently selected signal. The mean scores at baseline for signal items indicated that patients selected items for which the severity was intermediate, and which could therefore potentially show response (i.e., could either improve or deteriorate). Six of 20 patients (30%) selected as signals items for which they scored the pain as extreme (i.e., a score of 4); none selected signals scored as 0. The pain signal was usually ranked highest in severity, though it was often tied in severity with other nonsignal items.

Using both the signal and the aggregate strategies, there was statistically significant improvement at 6 months postsurgery versus baseline ($P < 0.001$), irrespective of the type of analysis (i.e., parametric or nonparametric). The relative efficiency (signal versus aggregate) was 1.00 for nonparametric analysis and 1.30 for parametric analysis. When individual items were analyzed, all 5 pain items detected statistically significant improvement in the pain level ($P \leq 0.004$). Clinical deterioration in those items not selected as signals (nonsignal deterioration) occurred on 4 occasions (i.e., 5% of item selections) in 3 patients. The magnitude of the deterioration was as follows: mean 1.00, SD 0, range 0. Sample size requirements were lower for the signal strategy (n = 17) than for the aggregate strategy (n = 30).

Stiffness. Patients designated both stiffness items as signals, with morning stiffness being selected slightly more frequently than "gelling" (Table 2). The mean scores at baseline suggested that patients were selecting potentially responsive signals. Two of 27 patients (7%) selected items for which the severity was rated as extreme; in both cases, this was the morning stiffness item. No patient selected as a signal an item for which the severity was scored as 0. The stiffness signal was often ranked highest, except in 2 instances, although it was tied in severity with the other nonsignal item.

Use of both the signal and the aggregate strategies enabled detection of statistically significant improvement ($P < 0.001$), regardless of the type of analysis. The techniques used were of similar relative efficiency (signal versus aggregate) for parametric analysis (RE = 1.00); the signal technique was slightly less efficient by nonparametric analysis (RE = 0.94). When the individual items were analyzed, both stiffness items detected statistically significant improvement ($P < 0.001$). Clinical deterioration in nonsignal items occurred in only 2 situations in 2 patients (i.e., 7% of item selections). The magnitude of the deterioration was as follows: mean 1.00, SD 0, range 0. Although the required sample sizes for the 2 strategies were quite similar, that for the aggregate strategy (n = 23) was lower than that for the signal strategy (n = 26).

Physical function. Eleven of the 18 physical function items in the WOMAC inventory were se-

Table 2. Stiffness dimension signals*

	No. of patients selecting item as a signal	Item ranking†	Signal score at baseline		Score at baseline		Change from baseline to 6 months (P)	
			Mean‡	SD	Mean‡§	SD	t-test (parametric)	Wilcoxon test (nonparametric)
Individual item								
Morning stiffness	16	1, 1, 1, 1¶, 1¶, 1¶ 1¶, 1¶, 1, 1, 1¶ 1, 1¶, 1, 1, 1¶	2.5	1.0	2.4	0.9	<0.001	<0.001
"Gelling"	11	1¶, 1, 1¶, 1, 1, 1¶ 1¶, 1¶, 1¶, 2, 2	2.1	0.7	2.0	0.8	<0.001	<0.001
Signal (n = 27)	NA	NA	2.3	0.9	2.3	0.9	<0.001	<0.001
Aggregate (n = 27)	NA	NA	4.4	1.6	4.4	1.6	<0.001	<0.001

* Morning stiffness = stiffness on first awakening; "gelling" = stiffness after sitting, lying, or resting later in the day; NA = not applicable.
† Item rank for each patient who selected item as a signal (possible number of ranks = 2).
‡ Except for the aggregate, possible scores could range from 0 (not affected) to 4 (most severe).
§ Derived by aggregating all scores at baseline for each individual item.
¶ Tied values. See Table 1 for explanation of ranking rules.

lected as signals (Table 3). Difficulty ascending stairs was the most frequently selected signal. The mean scores at baseline for signal items suggested that patients were selecting items for which there was potential response. Six of 24 patients (25%) selected items for which the degree of difficulty was rated as extreme (i.e., a score of 4); none selected a signal scored as 0. Rankings of the physical function signals selected varied from first to sixteenth, but they were usually ranked first, second, or third in severity.

Table 3. Physical function dimension signals*

	No. of patients selecting item as a signal	Item ranking†	Signal score at baseline		Score at baseline		Change from baseline to 6 months (P)	
			Mean‡	SD	Mean‡§	SD	t-test (parametric)	Wilcoxon test (nonparametric)
Individual item								
Descending stairs	2	1¶, 1¶	3.5	0.7	2.7	1.0	<0.001	0.001
Ascending stairs	5	2¶, 2¶, 1¶, 6¶, 1¶	3.2	0.8	2.9	1.0	<0.001	<0.001
Rising from sitting	0	NA	NA	NA	2.8	0.9	<0.001	<0.001
Standing	0	NA	NA	NA	2.4	1.1	<0.001	<0.001
Bending to floor	2	6¶, 5¶	2.0	0.0	2.4	1.3	<0.001	<0.001
Walking on flat surface	3	1¶, 11¶, 16¶	3.3	0.6	2.0	1.1	<0.001	<0.001
Getting in/out of car	3	3¶, 3¶, 1¶	2.7	0.6	2.8	1.1	<0.001	0.001
Going shopping	3	7¶, 1¶, 1¶	2.7	0.6	2.9	0.8	<0.001	<0.001
Putting on socks	1	1¶	4.0	ND	2.5	1.4	<0.001	0.001
Rising from bed	1	13¶	2.0	ND	2.2	1.2	<0.001	<0.001
Taking off socks	0	NA	NA	NA	2.0	1.4	<0.001	0.001
Lying in bed	1	8¶	3.0	ND	1.5	1.3	<0.001	0.002
Getting in/out of bath	1	5¶	2.0	ND	2.8	1.2	<0.001	<0.001
Sitting	0	NA	NA	NA	1.6	1.1	<0.001	0.001
Getting on/off toilet	0	NA	NA	NA	2.1	1.3	<0.001	<0.001
Heavy domestic duties	2	1¶, 3¶	3.5	0.7	3.3	1.2	<0.001	0.002
Light domestic duties	0	NA	NA	NA	1.8	1.2	<0.001	<0.001
Getting on/off a bus	0	NA	NA	NA	2.1	1.7	0.003	0.009
Signal (n = 24)	NA	NA	3.0	0.8	3.0	0.8	<0.001	<0.001
Aggregate (n = 24)	NA	NA	42.8	12.8	42.8	12.8	<0.001	<0.001

* NA = not applicable; ND = not determined (due to lack of sufficient numbers).
† Item rank for each patient who selected item as a signal (possible number of ranks = 18).
‡ Except for the aggregate, possible scores could range from 0 (not affected) to 4 (most difficult).
§ Derived by aggregating all scores at baseline for each individual item.
¶ Tied values. See Table 1 for explanation of ranking rules.

Both signal and aggregate strategies detected statistically significant improvement ($P < 0.001$), irrespective of the type of analysis. The signal strategy was slightly more efficient by parametric analysis (RE = 1.02) but slightly less efficient by nonparametric analysis (RE = 0.96). When individual items were analyzed, all physical function items detected statistically significant improvement, regardless of the type of analysis ($P \leq 0.002$). Clinical deterioration in nonsignal items occurred on 27 occasions in 7 patients (i.e., 7% of item selections). The magnitude of the deterioration was as follows: mean 1.56, SD 0.70, range 3. The sample size requirements were lower for the signal strategy (n = 12) than for the aggregate strategy (n = 16).

DISCUSSION

The principal objective of evaluative research is to detect clinically important and statistically significant alterations in health status. This objective can be most readily achieved by the application of highly responsive instruments to probe aspects of disease that are of defined importance. We have developed and reported on such an instrument (5,8), which is of potential value in the self-assessment of patients with OA of the hip or knee who have had surgical or pharmacologic intervention. To further enhance the statistical efficiency of the WOMAC Index and more closely tailor the measurement process to the unique symptom profile of the individual patient, we investigated the relative merits of using a single question from each of the 3 dimensions in the WOMAC inventory (signal technique).

To be considered an ideal alternative, the signal strategy would have to satisfy the following requirements: 1) Patients would differ in their selection of signal items, such that a single fixed item would not adequately convey the nature of all patients' symptomatology. 2) Patients would avoid signal items that lack response potential. This would certainly include the avoidance of items given a severity score of 0 and would likely entail the use of few, if any, items given a severity score of 4. 3) The signal strategy would be capable of detecting statistically significant ($P \leq 0.05$) alterations in health status with conventional sample sizes. 4) Relative efficiency would be greater than unity for the technique to be judged more efficient than the aggregate technique. As a result of increased RE, sample size requirements would be lower for the signal strategy than for the aggregate strategy. 5) The signal technique would adequately capture the patients' symptoms, such that deterioration occurring among items not selected as signals would not be overlooked.

Of course, an index meeting all of these criteria is unusual. However, a signal strategy meeting most of these criteria might still provide a useful alternative to more traditional forms of measurement. Several common themes emerge from the present study. The selection of the majority of items in the WOMAC inventory suggests that different patients place importance on different symptoms. The use of only a single fixed item is, therefore, deemed inappropriate and justifies the use of either a signal or aggregate approach to measurement. Indeed, the patient global assessment, which is often recommended and used in clinical studies, may itself be the result of a process in which the patient selects, weights, and aggregates a limited number of symptoms into an overall score (i.e., a signal strategy). However, the issue of selection cannot be decided merely by adopting a policy of selecting the signal by the most severely affected item in each WOMAC dimension since, as indicated in Tables 1–3, 71% of the first-placed rankings were tied.

It is also important to note that not all items selected as signals were ranked first in severity, and some of those selected were tied in rank. It is of substantial advantage, then, to allow the patient to designate which item is most important and will therefore be the principal object of observation and therapy for that patient. The success of the signal strategy in the present study is likely due to the fact that the vast majority of patients selected potentially responsive items as signals: Very few selected items for which the severity was scored as 4, and none selected items scored as 0. However, as indicated by the rankings, patients tended to select more severely affected items as signals. In all 3 dimensions, the signal strategy was responsive to change, demonstrating P values ≤ 0.002 in spite of the relatively small sample sizes used (n = 20–27). It should be noted that such high levels of significance may be a reflection of the potency of the surgical intervention. However, we have also observed similar levels of significance in another validation study of the WOMAC, in the context of a double-blind, randomized controlled trial of 2 nonsteroidal antiinflammatory drugs (8). We do not regard the observed success in this study as being generalizable to other subscales containing components that do not have strong potential for improvement or deterioration. The pain, stiffness, and physical function dimensions of

WOMAC contain only items that have been evaluated and found to have high response potential (5,8).

Relative efficiency is a crude, albeit convenient, method of comparing the effect size of different instruments. As such, its value may be affected in different studies by the severity of involvement for individual items and the frequency with which they are selected as signals. The utility of the RE as a measure of statistical economy of one instrument over another is uncertain and merits further evaluation. Nevertheless, the 3 dimensions included in the final WOMAC Index showed RE values close to unity, especially when the preferred nonparametric comparisons were considered. It is possible that the greater response of signal variables is explained by the effects of regression to the mean or a limit effect. We think it is unlikely that the response is merely a statistical aberration due to repeated random sampling (i.e., regression to the mean), since the Index as a whole also reflected the benefits of surgery. However, the further elucidation of a regression effect would require a randomized, "placebo"-controlled study, a design that poses major practical and ethical problems in the surgical setting. A limit effect is also unlikely, since in many cases, patients selected signals that were either tied at rank 1 or of a lower rank. Although many of the signals selected were items for which severity was rated as extreme, we believe that the fact that not all were makes the signal measure different from the simple selection of items scored as extreme, even though the numbers may not be that different.

With respect to sample size requirements, we have demonstrated that the signal strategy may be less demanding. With the exception of the stiffness dimension, sample size requirements were lower with the signal strategy than with the aggregate strategy. This, however, does not indicate that the signal strategy is necessarily the technique of choice. If, indeed, comprehensiveness is the key objective, then regardless of the slightly larger sample size requirement, the aggregate strategy remains the preferred technique. Indeed, investigators are frequently faced with the dilemma of selecting between simple and complex measures, each having different sample size requirements. The consequence of applying standard measures by signal or traditional techniques has been reported for only a small number of musculoskeletal indices. We anticipate that the present data will allow potential users to select the preferred mode of administration for future studies using the WOMAC Index. Deterioration in items not selected as signals occurred in each of the 3 subscales. However, these deteriorations were infrequent (9%) and of a low order of magnitude (mean 1.47; SD 0.74).

These data indicate that, while the signal strategy does not fulfill all of the ideal criteria mentioned above, from a practical standpoint, such a strategy may nevertheless provide a useful alternative to aggregate techniques of measurement, at least with respect to the WOMAC Index. In particular, it allows (a) a broader-based item selection than the use of a single fixed item, (b) the detection of statistically significant alterations in health status, (c) higher levels of RE tha the majority of either the aggregate or individual-item approaches to outcome measurement, and (d) reduced sample size requirements for studies using WOMAC as the principal outcome measure. It is limited by (a) the tendency of patients to select more severely affected items and, in some instances, items of extreme severity lacking response potential, and (b) its inability to detect nonsignal deterioration. The generalizability of our observations to future applications of WOMAC or, indeed, to signal methods of administration of other health status instruments, is, by necessity, limited. We do, however, agree with Tugwell and colleagues (4,10) that the signal technique merits further evaluation, since in some circumstances, it may facilitate attainment of high levels of statistical significance and clinical relevance.

REFERENCES

1. Dixon JS, Hill J, Bird HA: Measurement of proximal inter-phalangeal joint circumference in rheumatoid arthritis: one joint or ten? Br J Rheumatol 26:123–125, 1987
2. Ward JR, Niethammer TA, Egger MJ: Can we just measure signal joints? Should ring size and walking time be analyzed only in selected patients, Controversies in the Clinical Evaluation of Analgesic—Anti-Inflammatory—Antirheumatic Drugs. Edited by HE Paulus, GE Ehrlich, E Lindenlaub. New York, FK Schattauer Verlag, 1981
3. Egger MJ, Huth DA, Ward JR, Reading JC, Williams HJ: Reduced joint count indices in the evaluation of rheumatoid arthritis. Arthritis Rheum 28:613–619, 1985
4. Tugwell P, Bombardier C, Buchanan WW, Goldsmith CH, Grace E, Hanna B: The MACTAR patient preference disability questionnaire: an individualized functional priority approach for assessing improvement in physical disability in clinical trials in rheumatoid arthritis. J Rheumatol 14:446–451, 1987
5. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt L: Validation study of WOMAC: a health status

instrument for measuring clinically-important patient-relevant outcomes following total hip or knee arthroplasty in osteoarthritis. J Orthop Rheumatol 1:95–108, 1988
6. Colton T: Statistics In Medicine, Boston, Little, Brown and Co., 1974
7. The SPSS-X Users Guide. Third edition. Chicago, SPSS Inc., 1988
8. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW: Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 15:1833–1840, 1988
9. Liang MH, Larson MG, Cullen KE, Schwartz JA: Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. Arthritis Rheum 28:542–547, 1985
10. Tugwell P, Bombardier C: Practical issues in clinical trials as applied to current state of the art in rheumatology trials. Br J Rheumatol 22 (suppl):9–13, 1983

A Comparative Study of Signal Versus Aggregate Methods of Outcome Measurement Based on the WOMAC Osteoarthritis Index

SUSAN BARR, NICHOLAS BELLAMY, W. WATSON BUCHANAN, ANDREW CHALMERS, PETER M. FORD, WALTER F. KEAN, GUNNAR R. KRAAG, ERIKA GERECZ-SIMON, and JANE CAMPBELL

ABSTRACT. Objective. To compare signal versus aggregate measurement strategies using the VA3.0S version of the Western Ontario and McMaster Universities (WOMAC) Osteoarthritis (OA) Index.

Methods. Seventy patients with OA of the knee were asked to identify a signal item for each of the 3 dimensions of the WOMAC OA Index at baseline and termination of a 12-week, double blind, randomized, controlled trial.

Results. The signal method detected statistically significant alterations in health status at relatively small sample sizes and with a relative efficiency close to or at unity. In addition to a low prevalence of deterioration in nonsignal items, we observed some inconsistency in signal selection.

Conclusion. Signal methods of measurement may provide an alternative approach to outcome measurement provided issues of nonsignal deterioration and the consistency of signal selection can be addressed. (*J Rheumatol 1994;21:2106–12*)

Key Indexing Terms:
CLINIMETRICS OSTEOARTHRITIS CLINICAL STUDIES

Signal measurement is a methodologic technique in which measurement of disease is based on well defined, individualized targets, e.g., one or a few selected joint(s) or symptom(s). This technique has 2 objectives: (1) to tailor the measurement process to the symptom profile of individual patients and (2) to improve the efficiency of the measurement process by excluding aspects of disease that lack response potential. In spite of the possible advantages of signal measurement, it is noteworthy that neither Ward, *et al*[1] nor Egger, *et al*[2] demonstrated any clear superiority of the technique over more traditional methods in patients with rheumatoid arthritis (RA). However, Dixon, *et al*[3] and Tugwell, *et al*[4] have reported successful use of signal techniques in RA clinical trials, although Meenan and Pincus[5] have expressed concern regarding this approach to measurement. In the only previous comparative study of signal versus aggregate methods of outcome measurement in osteoarthritis (OA) clinical trials, we observed that the signal method was capable of detecting statistically significant alterations in health status with a relative efficiency close to unity and attended by a low prevalence of deterioration in nonsignal items[6]. However, that study was conducted in a group of patients undergoing total joint arthroplasty, an intervention attended by relatively large changes and small variance (cf, pharmacologic interventional studies).

We have extended our previous work by undertaking a comparison of signal versus aggregate measurement strategies using the VA 3.0S version of the Western Ontario and McMaster Universities (WOMAC) Osteoarthritis Index[7,8] in the context of a double blind, randomized, controlled trial of tenoxicam (Mobiflex™) and diclofenac (Voltaren™). The results of the between drug comparison of efficacy and tolerability have been reported in a separate publication[9]. WOMAC is a tri-dimensional, self-administered questionnaire probing clinically important, patient relevant outcomes in patients with OA of the hip and/or knee. It is valid, reliable, and of demonstrated responsiveness in surgical[7], pharmacologic[8,10], and physiotherapy[11] intervention studies and is capable of detecting a clinically important and statistically significant difference between 2 nonsteroidal antiinflammatory drugs (NSAID)[10]. The VA3.0S version is the 100 mm visual analog scale (VAS) version of the index that utilizes the signal (S) strategy. The 3.0 indicates that this is the original 3-subscale version of the index. Similarly the LK3.0 version is the 5-point adjectival Likert

From the Department of Medicine, The University of Western Ontario, London; McMaster University, Hamilton; Queen's University, Kingston, ON; University of British Columbia, Vancouver, BC; and The University of Ottawa, ON, Canada.

Supported by a research grant from Hoffmann-La Roche Limited.

S. Barr, MD, MSc, Fellow in Rheumatology, University of Calgary; N. Bellamy, MD, MSc, FRCP (Edin & Glas), FACP, Professor of Medicine, Epidemiology, and Biostatistics, University of Western Ontario; W.W. Buchanan, MD, FRFPS, FRCP (Glas, Edin), FRCPC, FACP, Professor, Department of Medicine, McMaster University; A. Chalmers, MD, FRCPC, Associate Professor, Department of Medicine, University of British Columbia; P.M. Ford, MB, LRCP (Lond), MRCP (Lond), MRCS (Eng), FRCPC, Professor, Queen's University; W.F. Kean, MD (Glas), FRCP(C), FACP, Clinical Professor, McMaster University; G.R. Kraag, MD, FRCPC, Professor, University of Ottawa; E. Gerecz-Simon, MD, Co-Director, Musculoskeletal Pain Research, Service of Rheumatology, McMaster University; J. Campbell, BA, Research Assistant, Department of Medicine, University of Western Ontario.

Address reprint requests to Dr. N. Bellamy, Suite 303, Colb. 3, South Street Campus, Victoria Hospital, 375 South Street, London, ON, Canada N6A 4G5.

Submitted December 20, 1993 revision accepted April 8, 1994.

form of the original 3-subscale version of the index. WOMAC VA3.0S is a form of the index in which patients not only respond to each of the 24 component items (5 pain, 2 stiffness, 17 physical function), but also identify at the point of introduction of a new intervention, one pain, one stiffness, and one physical function item of importance; i.e., they hope it will be favorably influenced by the subsequent intervention. The 3 items selected represent the principal focus (i.e., signals) of subsequent measurement.

We specifically wished to examine whether there was any advantage in using the WOMAC question inventory as a means of identifying signal symptoms for each of the 3 dimensions, compared with administering the inventory in its entirety. To address this issue, we identified (a) whether the signal strategy could be successfully applied in outcome measurement in OA of the knee, (b) the nature and severity of symptoms in the WOMAC inventory that were selected as signals, (c) the relative efficiency of the 2 techniques, (d) the sample size implications of the 2 techniques, (e) whether deterioration in nonsignal items was overlooked using the signal strategy, and (f) whether signal selection was the same at the beginning and end of the trial.

MATERIALS AND METHODS

The principal purpose of the original double blind randomized controlled trial was to compare the efficacy and tolerability of tenoxicam and diclofenac[9]. The methodologic aspects of the study, including the research architecture, inclusion and exclusion criteria, outcome measurement techniques employed, and statistical analysis have been reported elsewhere[9]. For current purposes, only those aspects of the design relevant to the comparison of different modes of utilization of the WOMAC index will be described.

In brief, 100 consecutive consenting outpatients with primary OA of the knee, who attended 6 participating rheumatology centers across Canada, were entered in the study[9] (Tables 1 and 2). At the screening visit, a history and examination were performed, and patients selected their worst knee as the primary focus of measurement for future assessment in the study. After NSAID-free washout period of 3 to 7 days, patients were reassessed using a number of measures including WOMAC VA 3.0S (Week 0), and only those who had deteriorated symptomatically (i.e., whose disease flared) were randomized in a double blind manner to receive either tenoxicam (one 200 mg capsule plus 2 placebo capsules daily) or diclofenac (three 50 mg capsules daily). At the baseline visit, patients not only completed the full WOMAC questionnaire, but also selected their 3 signal items (primary signals); this process being repeated at termination (secondary signals). The exact instructions were as follows: "Now we would like you to think again about each of the aforementioned symptoms, which you have just rated. Then select one pain item, one stiffness item, and one physical function item that is most important to you, i.e., that you most hope the treatment you are about to receive will improve." Patients were subsequently assessed at 2, 4, and 12 weeks. To maintain blinding, the capsules employed were identical in appearance and were taken 3 times daily with meals. During the washout period and the subsequent 12-week active treatment phase, rescue analgesia with monitored quantities of acetaminophen was allowed when necessary (up to six 325 mg tablets daily). Compliance to analgesic and study medications was verified by pill counting.

The current analysis has been restricted to baseline (Week 0) and termination (Week 12) data derived from WOMAC VA 3.0S and was similar to that described[6]. We have pooled the data from the tenoxicam and diclofenac groups since there was no statistically significant difference between the groups or their therapeutic response in the original study[9]. The mean baseline and termination scores for all primary signal items of each dimension, as well as for each individual item (i.e., signal and nonsignal scores) were calculated. Aggregate values for each dimension were calculated as the sum of the response scores (sum) of the component items. For each dimension, the termination and baseline data were compared using both parametric (Student's paired t test) and nonparametric (Wilcoxon matched pairs signed rank test) statistics[12]. The analysis was, therefore, restricted to patients who had (a) completed the study, (b) completed baseline and termination questionnaires in their entirety, and (c) had identified signal items at baseline. The test statistics were calculated using the SPSS-X software program[13], and p values ≤ 0.05 were considered significant. Since this was an exploratory analysis, no correction was made for multiple comparisons. Parametric techniques may be applicable for certain ordinal level data; however, our data were generally not normally distributed, and therefore, the use of nonparametric techniques provided a more conservative estimate of statistical significance[12]. Since the results of the nonparametric and parametric analyses of the change between baseline and termination were the same, we have reported only the nonparametric results. Results of studies using the WOMAC Index support this contention but suggest that this does not make any important difference in data interpretation[7,8].

The relative ranking of each signal item among other items in the same dimension was determined to assess whether patients selected signals pertaining to those aspects of their physical condition that were most severely affected. Nonsignal deterioration was determined by comparing baseline and termination scores for each item not identified as a primary signal in the full WOMAC inventory, and by noting the frequency and magnitude of any deterioration[6]. Deterioration was arbitrarily defined as an increase above the baseline score by 10 mm or more. The corresponding signal scores were also examined for concurrent deterioration.

Relative efficiency (RE) for the parametric analysis was calculated according to the method adopted by Liang, *et al*[14], where RE (signal versus aggregate) $= (t_{signal}/t_{aggregate})^2$. For the nonparametric tests, RE (signal versus aggregate) $= (Z_{signal}/Z_{aggregate})^2$. Sample size requirements, based on matched pairs analysis, were calculated for both the signal and aggregate strategies for each dimension, using a parametric technique[6]. Calculations were based on the assumptions that p values <0.05 were significant, the power of the test was 90%, and the difference to be detected could be in either direction, i.e., a 2-tailed test of the null hypothesis. For each strategy, the sample size formula used was as follows: n matched pairs of observations $= ([Z_{0.05} + Z_{0.10}]\sigma/D)^2$, where σ = the standard deviation of differences and D = the decimal difference (from baseline) to be detected (i.e., 0.25 of mean)[12].

RESULTS

Of the 100 patients entered in the study, 2 were excluded because of protocol violations (age >75, abnormal biochemistry results) and another 2 were excluded because of allergy to acetylsalicylic acid or acetaminophen. An additional 26 patients were not included in this analysis because they either did not select a primary signal (n = 6), or complete the WOMAC questionnaire in its entirety (n = 6), or because they withdrew from the study early (n = 14). Among the remaining 70 patients that were included in the present analysis, all selected a primary signal, but only 56 selected a secondary signal at study termination.

Pain. A statistically significant improvement in pain severity (termination versus baseline) was detected using both signal ($p < 0.001$) and aggregate ($p < 0.001$) measurement strategies. Pain scores improved significantly ($p \leq 0.02$) in each of the 5 individual items except for "pain while sitting or

lying" ($p = 0.41$). The percentage improvement in mean values was 30.5% for the signal and 21.7% for the aggregate technique. The signal strategy was relatively more efficient than the aggregate method based on both parametric (RE = 2.52) and nonparametric (RE = 1.91) analyses. Estimated sample size requirements were lower for the signal strategy (n = 25) than for the aggregate strategy (n = 30).

Each of the 5 items of the pain dimension were selected as signals by 2 or more patients, both at baseline and at study termination (Table 1). Patients were more likely to choose signal items probing pain severity during movement (baseline 71.4%, termination 76.8%) than at rest. The most frequently selected signal at baseline was "going up or down stairs" (47.1%) and at termination "walking on the flat" (42.9%). Primary and secondary signals were ranked as the most severely affected item in 43 and 23% of cases, respectively (Table 2). Only 10% of primary signal scores and 5.4% of secondary signal scores exhibited limited response potential related to extreme severity (i.e., ≥ 90 mm on a 100 mm VAS). None of the primary pain signals scored ≤ 10 mm, while 8.9% of secondary signal scores were ≤ 10 mm.

Stiffness. A statistically significant improvement in stiffness (termination versus baseline) was detected using both signal ($p < 0.001$) and aggregate ($p < 0.001$) measurement strategies. Stiffness scores improved significantly ($p < 0.001$) in each of the 2 individual items. The percentage improvement in mean values was 35.9% for the signal and 31.9% for the aggregate technique. The signal strategy was relatively more efficient than the aggregate method based on both parametric (RE = 1.39) and nonparametric (RE = 1.23) analyses. Estimated sample size requirements were lower for the signal strategy (n = 38) than for the aggregate strategy (n = 44).

Morning stiffness was selected as the signal slightly more often than "gelling" at baseline (58.6%) and at termination (51.8%) (Table 3). Primary and secondary signals were ranked as the most severely affected item in 73 and 82% of cases, respectively (Table 2). Primary and secondary stiffness signals were infrequently graded as being extreme in severity (5.7 and 1.8% scored ≥ 90 mm, respectively). Only 2.9% of primary stiffness signals and 32.1% of secondary signals had scores of ≤ 10 mm.

Physical function. A statistically significant improvement in physical function (termination versus baseline) was detected using both signal ($p < 0.001$) and aggregate ($p < 0.001$) measurement strategies. Scores of individual function items decreased significantly ($p < 0.001$) in 14 of the 17 individual items; items 9, 12, and 14 ($p = 0.001, 0.02, 0.003$, respectively). The percentage improvement in mean values was 32.3% for the signal and 24.9% for the aggregate technique. The signal strategy was relatively more efficient than the aggregate method based on both parametric (RE = 1.65) and nonparametric (RE = 1.15) analyses. Estimated sample size requirements were lower for the signal strategy (n = 28) than for the aggregate strategy (n = 40).

Of the 17 items comprising the function dimension, 14 were chosen as signals. Walking on the flat was the most frequently selected signal both at baseline (22.9%) and at

Table 1. *Pain dimension signal and aggregate scores*

Item	Number of Patients Selecting Item as Signal		Primary Signal Scores (mm)*				WOMAC Score (mm)* (n = 70)			
			Baseline		Termination		Baseline		Termination	
	Primary Signal	Secondary Signal	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1. Walking on a flat surface	17	24	46.0**	19.3	35.6**	24.3	43.1	20.8	35.8	25.1
2. Going up or down stairs	33	19	59.8	23.8	37.7	26.6	58.8	23.4	40.9	27.9
3. At night while in bed	13	8	62.5	18.3	44.6	30.7	39.9	27.1	31.9	28.3
4. Sitting or lying	2	2	62.0	8.5	61.5	41.7	33.0	21.4	30.6	27.3
5. Standing upright	5	3	63.2	22.0	46.6	17.6	47.3	24.7	34.8	25.1
Signal	70	56	57.3	21.9	39.8	26.4	N/A		N/A	
Aggregate (n = 70)	N/A	N/A	N/A	N/A	N/A	N/A	222.1	93.9	174.0	126.4

* Responses scored on 100 mm VAS. ** n is based on patients selecting as signal at baseline, i.e., primary signal.

Table 2. *Rank ordering of severity of selected signals at baseline (n = 70) and termination (n = 56)*

	Most Severe		2nd Most Severe		3rd Most Severe		4th Most Severe		Least Severe	
	Baseline	Termination	Baseline	Termination	Baseline	Termination	Baseline	Termination	Baseline	Termination
Pain	43*	23	33	34	10	27	10	4	4	12
Stiffness	73	82	N/A	N/A	N/A	N/A	N/A	N/A	27	18
Physical function	3	2	1	14	10	12	10	4	3	2

* The numbers are the percentage of patients ranking item in category.

Table 3. *Stiffness dimension signal and aggregate scores*

Item	Number of Patients Selecting Item as Signal		Primary Signal Scores (mm)*				WOMAC Score (mm)*			
			Baseline		Termination		Baseline		Termination	
	Primary Signal	Secondary Signal	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1. Morning stiffness	41	29	50.5**	24.4	29.4**	24.4	46.7	26.3	29.8	26.2
2. Gelling	29	27	48.9	22.9	35.4	26.2	42.6	23.5	31.0	25.1
Signal	70	56	49.8	23.7	31.9	25.1	N/A		N/A	
Aggregate (n = 70)	N/A	N/A	N/A		N/A		89.3	45.7	60.8	49.8

* Responses scored on 100 mm VAS. ** n is based on patients selecting item as signal at baseline, i.e., primary signal.

termination (23.2%) (Table 4). Primary and secondary signals were ranked as the most severely affected item in 3 and 2% of cases, respectively (Table 2). Only 10% of primary signal scores and 1.8% of secondary signal scores were extreme, i.e., ≥ 90 mm. Only 1.4% of primary function signals scored ≤ 10 mm, while 10.7% of secondary signals scores were ≤ 10 mm. Few signal scores were extreme, i.e., ≥ 90 mm (primary 10.0%; secondary 1.8%) or ≤ 10 mm (primary 1.4%; secondary 10.7%).

Nonsignal deterioration. Within the pain dimension, the baseline scores of 68 nonsignal items (24.3%) from 28 patients had increased by 10 mm or more at study termination indicating deterioration (Table 5). Among these 28 patients, the signal score had also deteriorated in 10 patients (35.7%). However the pain signal scores of the remaining 18 patients (64.3%) had remained the same (9/18) or improved (9/18). The mean increase in score of the 68 items was 26.2 mm with a range of 10 to 64 mm.

Nonsignal deterioration occurred in 12 stiffness items (17.1%) and was associated with concurrent signal deterioration in 5 of the 12 patients (41.7%). The remaining signal items had either improved (41.7%) or remained unchanged (16.6%). On average, the 12 nonsignal items had deteriorated by 24.4 mm (range 10–41 mm) (Table 5).

Physical function had declined in 145 nonsignal item scores (12.9%) of 32 patients (Table 5). Only 5 patients (15.6%) indicated concurrent worsening of the signal function item. Among the remaining 27 patients (84.4%), the signal had remained the same (9/32) or improved (18/32). The average magnitude of deterioration was 23.9 mm (range 10–75 mm).

Signal stability. A secondary signal was selected by 56 of the 70 patients at study termination. Of these, a new signal (i.e. different from baseline) was chosen for the pain, stiffness, and function dimensions by 44.6, 16.1 and 46.4% of patients, respectively. The remaining patients chose identi-

Table 4. *Physical function dimension signal and aggregate scores*

Item	Number of Patients Selecting Item as Signal		Primary Signal Scores (mm)*				WOMAC Score (mm)*			
			Baseline		Termination		Baseline		Termination	
	Primary Signal	Secondary Signal	Mean**	SD	Mean**	SD	Mean	SD	Mean	SD
1. Descending stairs	6	5	63.0	22.7	35.0	30.2	52.4	25.5	35.7	26.0
2. Ascending stairs	7	6	61.9	28.3	47.3	31.8	51.0	27.8	38.5	27.9
3. Rising from sitting	4	4	68.8	18.8	53.0	6.2	47.8	24.7	35.3	25.3
4. Standing	3	3	55.7	38.8	46.3	35.0	45.2	27.3	31.9	25.1
5. Bending to floor	6	0	67.8	26.5	42.8	33.3	46.6	31.6	35.8	28.6
6. Walking on flat	16	13	46.2	24.3	35.2	28.7	41.2	24.0	32.6	24.4
7. Getting in/out of car	4	3	67.0	30.8	34.8	27.7	52.4	26.7	37.6	25.2
8. Going shopping	4	3	60.3	20.2	37.0	9.5	50.9	27.6	37.8	27.1
9. Putting on socks/stockings	2	2	45.5	7.8	29.5	16.3	40.6	27.9	32.5	26.7
10. Rising from bed	0	0	N/A	N/A	N/A	N/A	42.3	25.5	30.4	23.9
11. Taking off socks/stockings	1	1	96.0	N/A	93.0	N/A	38.9	26.5	29.7	25.9
12. Lying in bed	2	1	47.5	6.4	15.0	18.4	31.1	24.8	25.5	25.8
13. Getting in/out of bath	3	3	69.0	29.5	66.7	28.1	47.2	28.3	35.8	27.6
14. Sitting	0	1	N/A	N/A	N/A	N/A	31.0	20.1	26.9	25.2
15. Getting on/off toilet	1	4	30.0	N/A	18.0	N/A	40.4	25.5	31.0	24.3
16. Heavy domestic duties	11	7	63.5	16.7	35.8	23.4	61.5	25.9	43.9	29.6
17. Light domestic duties	0	0	N/A	N/A	N/A	N/A	36.5	22.5	27.6	23.0
Signal	70	56	58.9	24.0	39.9	27.0	N/A	N/A	N/A	N/A
Aggregate (n = 70)	N/A	N/A	N/A	N/A	N/A	N/A	756.9	369.8	568.3	405.4

* Responses scored on 100 mm VAS. ** n is based on patients selecting item as signal at baseline, i.e., primary signal.

Table 5. *Deterioration of nonsignal items*

	Nonsignal Deterioration		Signal Item at Termination		
	No. Items (%)	No. Patients	Worse (No. of Patients) (%)	Same (%)	Better (%)
Pain	68 (24.3)	28	10 (35.7)	9 (32.1)	9 (32.1)
Stiffness	12 (17.1)	12	5 (41.7)	2 (16.6)	5 (41.7)
Physical function	145 (12.9)	32	5 (15.6)	9 (28.1)	18 (56.3)

cal primary and secondary signals; the signal score having improved (pain = 58%, stiffness = 66%, function = 50%), remained unchanged (pain = 32%, stiffness = 23%, function = 37%), or worsened (pain = 10%, stiffness = 11%, function = 13%).

Among those patients who changed signals, the primary signal had improved in the majority of cases (pain = 76%, stiffness = 78%, function = 89%), but remained the same in 11 to 16% (pain = 16%, stiffness = 11%, function = 12%), and worsened in 8 to 11% (pain = 8%, stiffness = 11%, function = 0%). Comparison of the scores of secondary signals at termination versus baseline revealed that most of the newly selected signal items had improved (pain = 60%, stiffness = 33%, function = 77%), although 19 to 33% (pain = 24%, stiffness = 33%, function = 19%) remained unchanged and 4 to 33% (pain = 16%, stiffness = 33%, function = 4%) worsened during the study. Scores of the secondary signal were compared with those of the primary signal at both baseline and termination. At Week 0, the secondary signal tended to be the same as (pain = 56%, stiffness = 56%, function = 62%) or less severely affected (pain = 36%, stiffness = 44%, function = 35%) than the primary signal item; although 4 to 8% (pain = 8%, stiffness = 0%, function = 4%) were more severely affected. At termination, however, the secondary signal tended to be the same as (pain = 76%, stiffness = 78%, function = 77%) or more severely affected (pain = 20%, stiffness = 22%, function = 19%) than the primary signal selected at baseline; although 4% (pain = 4%, stiffness = 0%, function = 4%) were less severely affected.

DISCUSSION

A fundamental goal of evaluative research is to develop efficient outcome measurement techniques capable of detecting clinically important and statistically significant changes in the health status of patients exposed to different therapeutic interventions. In theory, signal methodology restricts the measurement process to clinically relevant and potentially responsive components of disease[15]. Signals may enhance the ability to detect changes in health status by minimizing the dilution of response that occurs when uninvolved or irreversibly affected items are evaluated. The resulting "noise reduction" effect may offer several advantages, including (1) increasing mean change scores and reducing variance, (2) increasing the efficiency of outcome measurement, and (3) reducing sample size requirements. However, 3 main concerns have been raised regarding signal techniques: (1) patients may select more severely affected (and potentially less responsive) items as signals; (2) signal techniques may overlook important information by failing to detect changes in the status of nonsignal items; and (3) it is not known how reliable or stable signal selections are over time.

The WOMAC Index is a responsive health status instrument for the evaluation of patients with OA of the knee or hip. In a report on patients undergoing total joint arthroplasty[6], we investigated whether there was any advantage in using the WOMAC question inventory for identifying signal symptoms in each of the 3 dimensions (signal technique), compared with administering the inventory in its entirety (aggregate technique)[6]. In our present study, we have performed a similar comparison to determine whether the signal strategy could be successfully applied in outcome measurement in patients with knee OA undergoing a pharmacologic intervention and have attempted to explore further the proposed advantages and disadvantages of signal methodology described above.

The signal strategy was applied successfully in our study. As reported[6], both the signal and aggregate techniques detected a statistically significant improvement in pain, stiffness, and function compared with pretreatment status. This difference was apparent even when the more conservative nonparametric statistics were used. The 2 NSAID were found to be similar in efficacy as reported in a separate publication[9]. The estimated percent improvement of the mean scores was greater for the signal technique than the aggregate in the pain, function, and stiffness dimensions, although the difference in the latter was rather small. This difference may result from bias in the signal measurement due to the patient's hope for, or focus on, a particular outcome. Patients will tend to identify those items as primary signals where they are highly impaired and expect to gain the most improvement. Naturally there will be more room for improvement, although this does not necessarily mean that these items are the most responsive. We would expect, as shown in this study, that signal methods have a higher percentage improvement score than aggregate techniques.

The relative efficiency statistic is one method of comparing the responsiveness of different measurement instruments; however, its utility remains to be fully evaluated. An RE value of 1.0 suggests equal efficiency of the instruments being

compared; however, there are no standards for defining the significance of values < 1.0 or > 1.0. For example, an RE value of 2.0 does not indicate that the numerator index is necessarily "twice" as efficient as the denominator index. Nevertheless, if the RE is > 1.0, the instrument in the numerator can be inferred to be the more responsive outcome measure requiring smaller sample sizes and/or capable of detecting smaller effect sizes than the instrument in the denominator. In the current analysis, the signal strategy was more responsive than the aggregate in all 3 symptom dimensions, with RE values ranging from 1.39 to 2.52 by parametric analysis and 1.15 to 1.91 by nonparametric analysis. Accordingly, the sample size requirements for the signal strategy were lower. The results of our previous comparison of signal versus aggregate techniques in patients undergoing total joint arthroplasty were similar, although the RE values reported here are higher. Sample size requirements in the surgical setting were lower, likely as a result of the greater potency, and, therefore, magnitude of response to this intervention.

Responsiveness is of paramount importance for evaluative measures[15]. The WOMAC Index contains 24 items that have been formerly evaluated and found to have high response potential[7,8]. It seems likely that signal and global strategies compare favorably with respect to responsiveness only in those situations where patients select potentially responsive signal items from the menu[6]. Our patients tended to select signal items that were ranked as relatively more severe compared with the scores of the remaining items at baseline, the rankings being based on absolute scores in mm. Although other nonsignal items had scores similar in severity to those of signal items (e.g., within 5 or 10 mm), we felt that combining scores into artificial rank groupings based on an arbitrary cutoff (e.g., $\pm$ 10 mm) might result in a loss of information. Despite a high severity ranking, signal items were evidently responsive as shown by their ability to detect statistically significant improvements in health status with pharmacologic therapy. Very few of the primary signals lacked the potential to detect deterioration because of extreme severity ($\geq$ 90 mm) or improvement because of relative noninvolvement ($\leq$ 10 mm).

To assess nonsignal deterioration over time, we arbitrarily defined deterioration *a priori* as an increase in nonsignal item scores by $\geq$ 10 mm at termination versus baseline. Although it is recognized that this definition is arbitrary, a 10 mm increment in score represents a 22.5% deterioration of the mean baseline score (45 mm) of all items in each dimension. This appears to be a reasonable definition since, as a rule, changes of 20 to 25% are generally considered to be clinically important[16], and therefore, the importance of nonsignal deterioration would not be greatly over or underestimated. Using this definition, an average of 18.1% of all nonsignal items had deteriorated with an average increase in score of 25 mm (range 10–75). In many cases there was no concurrent deterioration of the signal items, i.e., important information might have been overlooked if the signal technique had been used alone. Although the importance of this degree of nonsignal deterioration is not known, its occurrence is a concern. Furthermore, if comprehensiveness (construct validity) is a priority, then an aggregate strategy must be used. In fact, WOMAC VA3.0S (and LK3.0S) represents a hybrid approach that provides a priority based signal score in addition to the more comprehensive aggregate health status score and maximizes the strengths of both techniques, an approach that has been suggested by Meenan and Pincus[5].

With regard to signal stability, it is interesting that many patients changed their signal selections at study termination. Our study was not designed to address the reason(s) why patients changed signal selection. However, it is clear that health status changed during the study, and trends in the data suggest that patients may have selected secondary signal items that had not improved during the course of therapy and had become subjectively more important. While it is not surprising that signal selection varies as functional status changes, it is not clear whether signal selection would remain stable over time if function did not actually change.

In conclusion signal method of measurement may provide an alternative approach to outcome measurement provided issues of nonsignal deterioration and the consistency of signal selection can be addressed. We are concerned regarding dependency of the measurement process entirely on signal methodology. Although the potential for reducing sample size requirements is encouraging, we agree with Meenan and Pincus[5] that the relative strengths of comprehensive versus prioritized assessments require considerable additional study. This could be accomplished by studying more diverse patient groups and other prioritization procedures. At the present time, for evaluative purposes, we only recommend use of the full WOMAC LK3.0 or VA3.0 osteoarthritis indices.

REFERENCES

1. Ward JR, Niethammer TA, Egger MJ: Can we just measure signal joints? Should ring size and walking time be analyzed only in selected patients? In: Paulus HE, Erblick GE, Lindenlaub E, eds. *Controversies in the Clinical Evaluation of Analgesic-Anti-inflammatory-Anti-Rheumatic Drugs.* New York: FK Verlag, 1981:103–10.
2. Egger MJ, Huff DA, Ward JR, Reading JC, Williams HJ: Reduced joint count indices in the evaluation of rheumatoid arthritis. *Arthritis Rheum 1985;*28:613–9.
3. Dixon JS, Hill J, Bird HA: Measurement of proximal interphalangeal joints circumference in rheumatoid arthritis: one joint or ten? *Br J Rheumatol 1987;*26:123–5.
4. Tugwell P, Bombardier C, Buchanan WW, Goldsmith CH, Grace E, Hannah B: The MACTAR patient preference disability questionnaire: an individualized functional priority approach for assessing improvement in physical disability in clinical trials in rheumatoid arthritis. *J Rheumatol 1987;*14:446–51.
5. Meenan RF, Pincus T: Editorial: The status of patient status measures. *J Rheumatol 1987;*14:411–4.

6. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Duku E: Signal measurement strategies: are they feasible and do they offer any advantage in outcome measurement in osteoarthritis? *Arthritis Rheum 1990;*33:739-45.
7. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt L: Validation study of WOMAC: A health status instrument for measuring clinically-important patient-relevant outcomes following total hip or knee arthroplasty in osteoarthritis. *J Orthop Rheumatol 1988;*1:95-108.
8. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW: Validation study of WOMAC: A health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol 1988;*15:1833-40.
9. Bellamy N, Buchanan WW, Chalmers A, Ford PM, Kean WF, Kraag GK, Campbell J: A multicenter study of tenoxicam and voltaren in osteoarthritis of the knee. *J Rheumatol 1993;*20:999-1004.
10. Bellamy N, Kean WF, Gerez-Simon E, Campbell J, Buchanan WW: Double blind randomized controlled trial (RCT) of sodium meclofenamate (Meclomen) and diclofenac sodium (Voltaren). Postvalidation reapplication of the WOMAC Osteoarthritis Index. *J Rheumatol 1992;*19:153-9.
11. Young SL, Woodbury MG, Fryday-Field K, Donovan T, Bellamy N, Haddad R: Efficacy of the interferential current stimulation alone for pain reduction in patients with osteoarthritis of the knee: A randomized placebo controlled clinical trial (abstr). *Phys Ther 1991;*(suppl 6)71:R088.
12. Colton T: *Statistics in Medicine.* Boston: Little, Brown, 1974.
13. The SPSS-X Users Guide, 3rd ed. Chicago: SPSS, 1988.
14. Liang MH, Larson MG, Cullen KE, Schwartz JA: Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. *Arthritis Rheum 1985;*28:542-7.
15. Bellamy N: Responsiveness. In: *Musculoskeletal Clinical Metrology.* Lancaster: Kluwer Academic, 1993:31-5.
16. Bellamy N: Composite indices. In: *Musculoskeletal Clinical Metrology.* Lancaster: Kluwer Academic, 1993:135-46.

Inflammopharmacology. 1993;2:85–87.

LETTER TO THE EDITOR

A study of the time frame dependency of responses to the WOMAC Osteoarthritis Index

The last few years have seen the development of a number of fully validated high-performance health status instruments relevant to outcome assessment in musculoskeletal clinical trials of surgical, pharmacological and physiotherapeutic interventions. The measures, relevant to total joint arthroplasty, were reviewed in two previous publications [1,2]. It is of note that some of these measures assess symptoms over specific periods of time (time frames) while others do not specify the interval. For example, the Health Assessment Questionnaire (HAQ) [3] probes health status over the previous week, while the Arthritis Impact Measurement Scales (AIMS) [4] require consideration of symptoms in the previous one month. In a review of functional status measures, Bombardier and Tugwell [5] have drawn attention to the importance of the appropriateness of the time frame for measuring the response to an intervention. Our experience with the WOMAC Osteoarthritis Index, in a number of different studies [6–10], has indicated that the time frame over which patients are asked to consider their symptoms often needs to be adjusted to respect varying induction–response dynamics of different interventions. We are aware from our previous work asking patients about their symptoms at precisely defined points in time (i.e. real time) and then applying sophisticated analytic techniques, that it is possible to detect circadian variation in the symptoms of osteoarthritis [11]. Furthermore, there is general concern that altering the wording of a questionnaire may alter its performance characteristics. Given the aforementioned time dependency of symptoms and the necessity to justify any changes in questionnaire wording, we have performed a preliminary evaluation of the effects of altering the time frame over which symptoms are rated on WOMAC subscale scores. Nineteen patients with primary knee osteoarthritis, who were enrolled in a 6-week, double-blind, randomized controlled trial of Meclomen versus Voltaren, participated in this study. The results of the between-drug comparison have been reported elsewhere [12]. The patients considered in this analysis were not subject to any alterations in pharmacological intervention in the last 2 weeks of this 6-week study and could be considered in steady state. There were 8 males and 11 females of mean age 57.2 years (range 43–65 years) and mean disease duration 5.9 years (range 2 months to 20 years). At the final assessment each patient was asked to complete, in random order, three versions of WOMAC differing only in the time frame over which patients were asked to consider their symptoms. The three time frames were: previous 24 hours, previous 48 hours, and previous 2 weeks. A break of several minutes was taken between presentation of each questionnaire, the patient being blind to responses given on any preceeding completed questionnaires. Data were analysed using analysis of variance techniques, examining, in particular, the effects of time and order. Sine the data were collected by two different centres (Hamilton and London), they were examined for any centre effect. In fact, no order effects or centre effects were noted. No statistically significant

or clinically important time-dependent differences were noted in scores on the pain, stiffness or physical function subscales of the WOMAC Osteoarthritis Index (Table 1).

TABLE 1
Descriptive statistics by levels of time*

WOMAC subscale	Previous 24 h		Previous 48 h		2 weeks before	
	X	SD	X	SD	X	SD
Pain	158.11	151.61	163.90	152.07	168.26	137.90
Stiffess	66.05	62.59	67.00	62.47	70.11	55.23
Physical function	566.21	494.47	554.78	494.88	573.16	450.07

*X = mean, SD = standard deviation

Although limited by the relatively small number of patients, we have observed no time dependency of questionnaire responses over the 14-day period. Given the dynamic nature of musculoskeletal symptomatology, it might be anticipated that the longer the interval, the more likely time might be a factor. If the interval is very short, one may detect symptom anomalies peculiar to a particular day, i.e. a good day or a bad day. If the interval is too long, patients may fail to recall adequately their symptoms. It seems likely, in this study, that patients are capable of averaging their symptoms over a specified time frame, and that, in doing do, any circadian or between-day effects are smoothed out.

From the aforementioned data we feel justified in varying the time frame over which questions are asked (at least between 1 and 14 days) when using the WOMAC Osteoarthritis Index, depending on the dynamic requirements of the study. We are unaware of any similar data having been published on any other musculoskeletal health status instrument.

G. Griffiths[1], N. Bellamy[1]*, W.F. Kean[2], J. Campbell[1] and E. Gerecz-Simon[2]
[1]Department of Medicine, University of Western Ontario, London, Ontario, Canada; [2]Department of Medicine, McMaster University, Hamilton, Ontario, Canada
*Correspondence: Dr N. Bellamy, Suite 303, Colbourne 3, Victoria Hospital (S. St. Campus), 375 South Street, London, Ontario, Canada N6A 4G5

REFERENCES

1. Bellamy N, Campbell J. Hip and knee rating scales for total joint arthroplasty: a critical but constructive review. Part I. J Orthop Rheum. 1989;3:3–21.
2. Bellamy N, Campbell J. Hip and knee rating scales for total joint arthroplasty: a critical but constructive review. Part II. J Orthop Rheum. 1989;2:63–76.
3. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. Arthritis Rheum. 1980;23:137–45.

4. Meenan RF, Gertman PM, Mason JH. Measuring health status in arthritis. The Arthritis Impact Measurement Scales. Arthritis Rheum. 1980;23:146–52.
5. Bombardier C, Tugwell P. Methodological considerations in functional assessment. J Rheumatol. 1987;14(Suppl.15):6–10.
6. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt L. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes following total hip or knee arthroplasty in osteoarthritis. J Orthop Rheum. 1988;1:95–108.
7. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol. 1988;15:1833–40.
8. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Duku E. Signal measurement strategies: are they feasible and do they offer any advantage in outcome measurement in osteoarthritis? Arthritis Rheum. 1990;33:739–45.
9. Bellamy N, Goldsmith CH, Buchanan WW, Campbell J, Duku E. Prior score availability: observations using the WOMAC Osteoarthritis Index. Br J Rheumatol. 1991;30:150–1.
10. Bellamy N, Wells G, Campbell J. Relationship between severity and clinical importance of symptoms in osteoarthritis. Clin Rheumatol. 1991;10:138–43.
11. Bellamy N, Sothern RB, Campbell J. Rhythmic variations in pain perception in osteoarthritis of the knee. J Rheumatol. 1990;17:364–72.
12. Bellamy N, Kean WF, Buchanan WW, Gerecz-Simon E, Campbell J. Double blind randomized controlled trial of sodium meclofenamate (meclomen) and diclofenac sodium (voltaren): post validation reapplication of the WOMAC Osteoarthritis Index. J Rheumatol. 1991;19:153–9.

MS received: 27 October 1992
Accepted for publication: 14 November 1992

Inflammopharmacology. 1995;3:1–6.

A COMPARATIVE STUDY OF THE RELATIVE EFFICIENCY OF THE WOMAC, AIMS AND HAQ INSTRUMENTS IN EVALUATING THE OUTCOME OF TOTAL KNEE ARTHROPLASTY

G. GRIFFITHS[1], N. BELLAMY[1], W.H. BAILEY[2], S.I. BAILEY[2], A.C. McLAREN[2] AND J. CAMPBELL[1]
[1]Department of Medicine, [2]Orthopaedic Surgery, University of Western Ontario, London, Ontario, Canada
Correspondence: Dr N. Bellamy, Director, Rheumatic Disease Unit, Suite 303, Colbourne Building 3, Victoria Hospital, SSC, 375 South Street, London, ON N6A 4G5, Canada

ABSTRACT

Griffiths G, Bellamy N, Bailey WH, Bailey SI, McLaren AC, Campbell J. A comparative study of the relative efficiency of the WOMAC, AIMS and HAQ instruments in evaluating the outcome of total knee arthroplasty. Inflammopharmacology. 1995;3:1–6.

The relative efficiency of the WOMAC, AIMS, and HAQ instruments was compared in 21 patients with osteoarthritis who received total knee replacement surgery. Assessments of pain and physical function were made pre-operatively and 6 months post-operatively and the relative efficiency of the three instruments in detecting change was determined. Overall, the WOMAC Osteoarthritis Index was greater in relative efficiency than the HAQ and AIMS instruments, particularly in the detection of change in functional status. We conclude that the WOMAC Index can be used as an alternative to the HAQ or AIMS instruments in outcome assessment in osteoarthritis clinical trials.

Keywords: arthroplasty, health measures, osteoarthritis, relative efficiency

INTRODUCTION

Total joint arthroplasty of the hip and knee is one of the most important orthopaedic surgical advances of recent decades. As a consequence, evaluation of the results obtained using total joint arthroplasty is of extreme importance. The methods used have been recently reviewed [1,2] and show considerable variability. In addition to the so-called 'orthopaedic indices', there have emerged several 'rheumatologic indices', which may be suitable for evaluating the outcome of total joint arthroplasty [3]. Indeed, Liang et al. compared four such indices [Health Assessment Questionnaire (HAQ), Arthritis Impact Measurement Scale (AIMS), Index of Well Being (IWB) and Functional Status Index (FSI)] with respect to their relative statistical efficiency in patients undergoing total joint arthroplasty [4]. We have previously compared the relative efficiency of our own index (WOMAC) and the Lequesne Algofunctional Index and found them similar in relative efficiency [5]. WOMAC (Western Ontario McMaster Universities Osteoarthritis Index) is a 24-item self-administered questionnaire which separately probes the three dimensions of pain, stiffness and physical function. The reliability, validity, and responsiveness of both the visual analogue-scaled (VA3.0) and Likert-scaled (LK3.0) versions have been evaluated [5,6]

as have issues of parametric versus non-parametric analysis [5,6], prior score availability [7], weighting and aggregation [8], time frame dependency [9] and the responsiveness of WOMAC in surgical [5], pharmacologic [6,10,11] and physiotherapy [12] interventions. The favourable experience with WOMAC has prompted us to compare the relative efficiency of the WOMAC, HAQ [13], and AIMS [14] instruments in osteoarthritis patients undergoing total joint arthroplasty using the method described by Liang et al. [4].

METHODS

Twenty-six consenting patients with primary osteoarthritis (OA) scheduled for total knee arthroplasty at Victoria Hospital, London, Ontario, were chosen for study. Patients who had undergone prior total knee replacement surgery were excluded, as were patients with visual or hearing impairments, and those who could not read or understand English. The day prior to surgery one of the investigators (GG) personally interviewed each patient to collect the following pre-operative, baseline data: age, gender, disease duration, ARA functional class, joint geometry measures, radiographic grading (Kellgren and Lawrence method [15]), and Hospital for Special Surgery (HSS) Knee Rating Scale [16].

In addition, three health status instruments were completed pre-operatively (day prior to surgery), and 6 months post-operatively: the AIMS, HAQ and WOMAC (VA3.0). The AIMS is a multidimensional instrument probing physical, social and emotional well-being. Although the index has nine components, only the mobility, physical activity, dexterity, household activity, activities of daily living, and pain subscales were analysed; however, data were collected, but not reported, on the remaining 3 scales (social activity, depression, anxiety). The HAQ disability index assesses the patient's functional ability over the past week. For each of 8 categories (dressing and grooming, arising, eating, walking, hygiene, reach, grip and activities), the patient indicates the amount of difficulty experienced over the last week. The use of devices and/or help from another person is factored into the index. A 15-cm visual analogue scale is used to measure arthritis pain. The WOMAC OA Index (VA3.0) was administered with a one-week time frame, responses being made on 100 mm visual analogue scales. The 3 health status instruments were administered in random order to each patient pre-operatively and again by mail 6 months later. We had originally planned to obtain corresponding data on the HSS. However, while all patients agreed to be assessed prior to surgery, most were from outlying areas and did not wish to return to London for the 6-month post-operative clinical assessments.

Data analysis was performed using the SAS package of statistical programmes. Means and standard deviations were computed for pre-operative scores and 6-month post-operative scores. Relative efficiency (RE) was calculated using the method described by Liang et al. [4], e.g. RE for WOMAC versus HAQ = $(t_{WOMAC}/t_{HAQ})^2$. An RE >1 means that the WOMAC is more efficient than the HAQ; while RE <1 means that the WOMAC is less efficient than the HAQ.

RESULTS

Of the 26 consenting surgical candidates originally screened for this study, follow-up data were not obtained on 5 patients who failed to return their 6-month post-operative questionnaires. The demographic profile of the 21 patients analysed is indicated in Table 1. The mean age and disease duration were 65.1 years and 9.5 years, respectively. Most patients were female. The mean HSS score was 61. Most patients were of Radiographic Grade II or III, and ARA Functional Class II or III. No Radiographic Grade I or Functional Class I or IV patients were included in the study.

TABLE 1
Characteristics of patients ($n=21$)

		Mean	SD
Age (years)		65.1	8.4
Disease duration (years)		9.5	8.0
Flexion (degrees)		106.5	21.4
Extension (degrees)		5.2	6.3
Varus (degrees)		4.3	4.3
Valgus (degrees)		1.5	2.6
Hospital for Special Surgery Score		61.0	11.5
Gender (male:female)		7:14	
X-ray grading	I	$n=0$	
	II	$n=7$	
	III	$n=12$	
	IV	$n=2$	
ARA Functional Class	II	$n=16$	
	III	$n=5$	

Mean and standard deviation values for pre-operative and 6-month post-operative assessments on the AIMS, HAQ and WOMAC instruments are illustrated in Table 2. Mean scores for pain and function declined following surgery on all 3 instruments, all changes being statistically significant ($p<0.001$ to 0.004, Table 2).

The RE of the WOMAC versus the AIMS and HAQ instruments was based on t values obtained by Student's paired t-test using pre-operative and 6-month post-operative scores for pain and function. The RE of WOMAC versus AIMS was as follows: pain = 0.81, physical function = 1.75. The relative efficiency of WOMAC versus HAQ was as follows: pain = 1.59, physical function = 1.13.

TABLE 2
Pre-operative and 6-month post-operative scores for the three health status instruments

Outcome measure		AIMS			HAQ			WOMAC		
		Pre-op	6-mth	*p* value	Pre-op	6-mth	*p* value	Pre-op	6-mth	*p* value
Pain	mean	6.45	4.43	<0.001	1.83	1.11	0.004	248.25	127.70	0.001
	SD	1.92	2.18		0.81	0.83		126.58	84.22	
Function	mean	2.30	1.62	0.003	1.01	0.69	<0.001	876.70	506.70	<0.001
	SD	0.97	1.09		0.42	0.57		367.55	348.55	

DISCUSSION AND CONCLUSIONS

The decision of which instrument to use in outcome measurement is based on a number of criteria, particularly validity, reliability and responsiveness. The clinimetric properties of the WOMAC, HAQ and AIMS instruments have been well established and meet acceptable standards [3]. Given that all three are high performance instruments, they are differentiated mainly by conceptual and statistical issues. The HAQ and AIMS instruments were principally developed for RA patients, although they have subsequently been validated in OA and mixed arthritis populations. In contrast, the WOMAC is a purpose-built instrument based on comprehensive interviews of 100 OA hip and knee patients. The HAQ probes many upper extremity functions that are not relevant when assessing the symptomatology of patients with OA of the hip and knee. In contrast the AIMS is a more comprehensive measure of health status in arthritis patients but probes aspects of social and emotional functioning in addition to pain and physical disability. In contrast, WOMAC focuses exclusively on the discomfort and disability associated with OA of the hip and knee. Furthermore, the WOMAC is a shorter instrument requiring responses to only 22 pain and physical function questions (cf HAQ = 44, AIMS = 29). There are, therefore, conceptual reasons to prefer the WOMAC instrument in situations when the goal is specifically to assess change in pain and physical disability due to OA of the hip and/or knee. Despite such advantages, a purpose-built instrument might be less desirable if the statistical efficiency with which it detected change was less than that of competing alternatives. The relative efficiency statistic is one method of comparing instruments. In this study, the WOMAC pain scale was slightly less efficient than that of the AIMS but more efficient than that of the HAQ. However, the WOMAC physical function scale was more efficient than that of the AIMS, and slightly more efficient than that of the HAQ. Given that RE values of WOMAC were >1 in 3 out of 4 comparisons, we conclude that overall the WOMAC is more efficient. Thus, for conceptual and statistical reasons the WOMAC offers some advantage over competing alternatives.

This study involved a spectrum of patients varying in age, gender, disease duration, Functional Class and Radiographic Grade. Total joint arthroplasty is a potent intervention in which changes are relatively large and the variance relatively small even in such diverse patients. This explains why very small p values were obtained despite a sample size of only 21 patients. However, as Liang et al. have noted 'The issue (sic, relative efficiency) is critical in situations where the expected changes are smaller than those seen in joint replacement, settings that include medical (sic, pharmacologic) and rehabilitative management of arthritis'. To date there are no international standards of measurement for OA clinical trials of anti-rheumatic drugs. Indeed, current Food and Drug Administration [17] and European League Against Rheumatism [18] guidelines for NSAID studies are not in complete agreement. At the Fifth Joint World Health Organization/International League Against Rheumatism Task Force Meeting on Rheumatic Diseases in Geneva, a proposal for standardization made at the First Osteoarthritis Research Society (OARS) Congress was discussed. The WOMAC index has been proposed as a potentially suitable outcome measure to assess pain and function in trials of slow-acting drugs in OA [19]. To date we have received requests from 89 investigators in 14 different countries for permission to use the WOMAC index in their studies. Given its widespread use, the

results of this study support the contention that in the assessment of pain and physical disability in OA hip and knee studies, the WOMAC index offers some conceptual and statistical advantages over alternative instruments.

REFERENCES

1. Bellamy N, Campbell J. Hip and knee rating scales for total joint arthroplasty: A critical but constructive review – Part I. J Orth Rheumatol. 1989;2:3–21.
2. Bellamy N, Campbell J. Hip and knee rating scales for total joint arthroplasty: A critical but constructive review – Part II. J Orth Rheumatol. 1989;2:63–76.
3. Bellamy N. Health status instruments and functional indices. In: Musculoskeletal Clinical Metrology. Lancaster: Kluwer Academic Publishers; 1993:77–101.
4. Liang MH, Larson MG, Cullen KE, Schwartz JA. Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. Arthritis Rheum. 1985;28:542–7.
5. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: A health status instrument for measuring clinically important patient relevant outcomes to anti-rheumatic drug therapy in patients with osteoarthritis of the hip and knee. J Rheumatol. 1988;15:1833–40.
6. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt L. Validation study of WOMAC: A health status instrument for measuring clinically-important patient-relevant outcomes following total hip or knee arthroplasty in osteoarthritis. J Orth Rheumatol. 1988;1:95–108.
7. Bellamy N, Goldsmith CH, Buchanan WW, Campbell J, Duku E. Prior score availability: observations using the WOMAC Osteoarthritis Index. Br J Rheumatol. 1991;30:150–60.
8. Stewart AL, Ware JE, Brook RH. Advances in the measurement of functional status: construction of aggregate indexes. Medical Care 1981;XIX(5):473–88.
9. Griffiths G, Bellamy N, Kean WF, Campbell J, Gerecz-Simon E. A study of the time frame dependency of responses to the WOMAC Osteoarthritis Index. Inflammopharmacology. 1993;2:85–7.
10. Bellamy N, Kean WF, Buchanan WW, Gerecz-Simon E, Campbell J. Double-blind randomized controlled trial of sodium meclofenamate (Meclomen) and diclofenac sodium (Voltaren): Post validation reapplication of the WOMAC Osteoarthritis Index. J Rheumatol. 1992;19:153–9.
11. Bellamy N, Buchanan WW, Chalmers A, Ford PM, Kean WF, Kraag GR, Gerecz-Simon E, Campbell J. A multicenter study of tenoxicam and diclofenac in patients with osteoarthritis of the knee. J Rheumatol. 1993;20:999–1004.
12. Young SL, Woodbury MG, Fryday-Field K, Donovan T, Bellamy N, Haddad R. Efficacy of interferential current stimulation alone for pain reduction in patients with osteoarthritis of the knee: A randomized placebo control clinical trial. (Abstract). Physiotherapy Canada. 1991;42(2):4.
13. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. Arthritis Rheum. 1980;23:137–45.
14. Meenan RF, Gertman PM, Mason JM. Measuring health status in arthritis: The Arthritis Impact Measurement Scales. Arthritis Rheum. 1980;23:146–52.
15. Atlas of Standard Radiographs – The Epidemiology of Chronic Rheumatism. Vol. 2. Philadelphia: FA Davis; 1963.
16. Insall JN, Ranawat CS, Aglietti P, Shine J. A comparison of four models of total knee-replacement protheses. J Bone Joint Surg. 1976;58-A-6:754–65.
17. Guidelines for the clinical evaluation of anti-inflammatory and anti-rheumatic drugs (adults and children). US Department of Health and Human Services. Public Health Services, Food and Drug Administration. April 1988:12–5.
18. Guidelines for the clinical investigation of drugs used in rheumatic diseases, European Drug Guidelines Series 5, World Health Organization, Regional Office for Europe, Copenhagen, European League Against Rheumatism, March 1985:21–4.
19. Lequesne M, Brandt K, Bellamy N, Moskowitz R, Menkes J, Pelletier JP, Altman R. Guidelines for testing slow acting drugs in osteoarthritis. J Rheumatol. 1994;(Suppl 41)21:65–71.

Manuscript received 17 May 94.
Accepted for publication 8 July 94.

Comparison of the Responsiveness and Relative Effect Size of the Western Ontario and McMaster Universities Osteoarthritis Index and the Short-Form Medical Outcomes Study Survey in a Randomized, Clinical Trial of Osteoarthritis Patients

Glenn M. Davies, Douglas J. Watson, and Nicholas Bellamy

Objective. *This study compares the responsiveness and relative effect sizes of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) with the Medical Outcomes Study Short Form Health Survey (SF-36) in a randomized clinical trial for treatment of osteoarthritis (OA).*

Methods. *Patients with OA of the knee or hip were randomized to receive either placebo or 2,400 mg/day of ibuprofen for 28 days. Patients completed the WOMAC and SF-36 at baseline and days 7, 14, and 28 of the trial.*

Results. *Patients receiving ibuprofen showed significant improvement in WOMAC pain, physical functioning, and the total score, while improvement was detected only for bodily pain on the SF-36. The WOMAC detected significant differences between ibuprofen and placebo for pain and physical functioning, whereas the SF-36 detected differences for the bodily pain subscale.*

Conclusion. *These results suggest the WOMAC has greater power to detect treatment differences than the SF-36, with respect to pain and physical functioning, in OA clinical trials.*

Key words. *Osteoarthritis; Western Ontario and McMaster Universities Osteoarthritis Index; Medical Outcomes Study Short Form Health Survey; Responsiveness; Relative effect size.*

INTRODUCTION

Therapeutic drug trials for osteoarthritis (OA) typically involve a number of efficacy measures, including both investigator and patient assessments of disease status, disease activity, and response to therapy. Usually, patient self-assessments will include measures of pain and stiffness in affected joint(s), and physical function or disability. These measures must be valid, reliable, and responsive to clinically meaningful change in order to differentiate between treatments (1). Several measures originally developed for rheumatoid arthritis have been adapted for use in clinical studies of OA. These include the Arthritis Impact Measurement Scales (1), the Health Assessment Questionnaire (HAQ) (2), and the Doyle Index (3). OA-specific measures include the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) (4) and the Lequesne Index (5). The HAQ

Glenn M. Davies, DrPH, and Douglas J. Watson, PhD, Merck Research Laboratories, West Point, Pennsylvania; and Nicholas Bellamy, MD, MS, Division of Rheumatology, Victoria Hospital, London, Ontario, Canada.

Address correspondence to Glenn M. Davies, DrPH, Department of Epidemiology, Merck Research Laboratories, PO Box 4 (BL 1-7), West Point, PA 19486-0004.

Submitted for publication September 10, 1998; accepted in revised form December 8, 1998.

0893-7524/99/$5.00

has been shown to be as sensitive as observer-based measures in clinical trials for OA (6) and was able to differentiate OA patients from individuals with other rheumatic diseases (7).

An alternative to disease-specific measures are generic instruments. The Medical Outcomes Study 36-item Short Form Health Survey (SF-36) is one such general health status instrument; it includes bodily pain and physical function scales as well as scales that evaluate social, mental, and emotional constructs. This instrument has been used in conjunction with the WOMAC to evaluate the health status of patients after total knee replacement (8,9).

An instrument that efficiently detects response to effective therapy can significantly reduce the number of patients necessary to detect meaningful treatment differences and allow better clinical decisions with respect to efficacy. This study compares the responsiveness and relative effect size of the WOMAC with that of the SF-36 in a randomized, controlled trial of 2,400 mg/day of ibuprofen versus placebo in patients with OA of the hip or knee. Specifically, scores on the WOMAC pain and physical function scales are compared with those of the SF-36 bodily pain and physical function scales.

PATIENTS AND METHODS

The trial was a multicenter, double-blind, parallel group clinical trial of patients with American College of Rheumatology (ACR; formerly the American Rheumatism Association) functional class I–III OA of the hip or knee for greater than 6 months (10,11). Classification of functional status was based on the revised criteria of the ACR. Subjects were required to have a well-established diagnosis of OA of the hip or knee, manifested by pain in the affected joint on motion or weight bearing for the majority of days during the month prior to study entry. Radiographs were not required as part of the diagnosis of OA. The diagnoses were made by primary care physicians from 20 different clinics across the continental United States.

Patients currently taking nonsteroidal anti-inflammatory drugs (NSAIDs) for pain relief were only entered into the study if they were taking short-acting formulations. Patients were excluded if they were mentally or legally incapacitated, had a history of a serious adverse event related to NSAID use, had a history of gastric or duodenal ulcer, or showed evidence of gastrointestinal (GI) bleeding on stool hemoccult. Patients with gastroesophageal reflux disease; disease of the esophagus, stomach, liver, gallbladder, pancreas, or small or large bowel; or other illness that, in the opinion of the investigators, might pose an additional risk to the patient or confound assessment of GI symptoms related to the use of the test treatment were excluded. Patients were also excluded if they were taking methotrexate or oral or systemic corticosteroids.

Following an initial 4-day placebo run-in period to assess adherence, subjects who met the inclusion/exclusion criteria were randomized 1:1 to receive either placebo or 800 mg of ibuprofen 3 times daily (2,400 mg/day) for 4 weeks. Both treatment groups were allowed to take 650 mg of open-label acetaminophen 4 times daily as needed for pain during the run-in and the 4-week treatment. Patients self-administered the 48-hour visual analog scale (VAS) version 3.0 of the WOMAC, the 1-week (acute) version of the SF-36, and a global satisfaction with medication question at baseline and days 7, 14, 17, and 28. Standard patient and laboratory exams and safety assessments were conducted at baseline and during treatment, and monitoring for adverse experiences was done at each visit. Use of concomitant medications, including acetaminophen, was recorded during the course of the trial. Patients were discontinued from the study for lack of efficacy, side effects/toxicity, or other reasons. Patients gave informed consent before participating in the study, and the protocol was approved by an institutional review board.

Baseline demographic statistics were computed overall and by treatment group for age, race, sex, ACR functional class, current NSAID use, duration of arthritis, and location of OA. Mean, standard deviation, minimum, maximum, and median were computed for continuous variables, and counts and percentages were computed for categoric variables.

All scales of the WOMAC (12), the SF-36 (13), and global satisfaction were scored on a 0–100 scale. In addition to the WOMAC scale scores, a WOMAC total score based on importance-weighting was also computed (12). WOMAC scale scores and the global satisfaction score were then reverse-scored by taking 100 minus the original score. In addition to the 8 SF-36 scale scores, two component summary scores, the physical component summary and the mental component summary, were computed (14). Higher scores indicate a better response to treatment on all measures. Detailed scoring information for the WOMAC (12) and the SF-36 (13) have been published previously.

Day 14 scores for all 8 scales and the component summary scores of the SF-36 were correlated, using

Spearman rank correlation coefficients, with the 3 WOMAC scale scores and the total WOMAC score to assess congruent validity (15). Day 14 change scores for the two instruments, and for the global satisfaction with medication question, were also correlated using the same method to assess longitudinal congruent validity (15).

Analysis of variance (ANOVA) methods were used to evaluate the within-treatment differences (responsiveness) and the between-treatment differences (relative effect size) of the SF-36, WOMAC, and global satisfaction. Change scores were computed for days 7, 14, and 28 for each of the scale scores as well as the component summary and total scores. Mean change scores adjusted for center effects were computed for each treatment group using SAS PROC GLM (16). One-sample *t*-tests were used to assess within-treatment group changes, and 2-sample *t*-tests were used to assess between-treatment group changes. Unadjusted mean change scores and standard deviations of these changes were computed for each treatment group. Effect sizes were computed by dividing the estimated mean difference of scores between groups by the pooled standard deviation (mean square error) from the ANOVA model.

RESULTS

The study enrolled 104 OA patients, 50 to placebo and 54 to ibuprofen. Descriptive statistics, overall and by treatment group, are given in Table 1. The mean age of the sample subjects was 61.5 years, and the sample was primarily white (83.7%), well educated (92.3% with education > high school), and female (63.5%). With respect to OA disease history, the majority of patients had OA of the knee (68.3%), were either ACR functional class I or II (86.5%), and had an average duration of OA of 7.9 years. Approximately 90% of the sample were taking short-acting NSAIDs prior to entering the study. There were no significant differences between treatment groups with respect to these baseline characteristics.

Table 2 shows the means and standard deviations for scale scores over time for each treatment group. The ibuprofen group showed improvement in all WOMAC scale scores within the first week. These improvements were maintained for the duration of the study. The ibuprofen group also showed improvement in SF-36 pain scores within the first week that lasted for the duration of the study; improvements in SF-36 physical functioning and role physical scores occurred within two weeks of study start, but were of a lesser magnitude and attenuated somewhat over time. Moderate improvements in the SF-36 energy/vitality score were also seen with ibuprofen. No other changes occurred in the ibuprofen group with the SF-36 scales during the 28 days of the study. In the placebo group, no notable changes were seen over time on either the WOMAC or the SF-36, with the exception of a modest increase in scores on the role mental scale.

Analyses to assess congruent validity showed strong correlation between the pooled treatment WOMAC pain and physical function, and the SF-36 bodily pain and physical function scale scores (Figure 1). The correlation between day 14 scale scores was highly significant: r = 0.64 for WOMAC pain and SF-36 bodily pain, r = 0.72 for WOMAC physical function and SF-36 physical function. The correlation coefficients for the individual treatment

Table 1. Study demographics*

Variable	Placebo (n = 50)	Ibuprofen (n = 54)	Total (n = 104)
Age, years			
Mean	62.1	61.0	61.5
SD	7.2	9.3	8.2
Minimum	45.0	45.0	45.0
Median	61.5	61.0	61.1
Maximum	79	77	79
Race, %			
Black	12.0	13.0	12.5
White	82.0	85.2	83.7
Other	6.0	1.8	3.8
Sex, %			
Men	36.0	37.0	36.5
Women	64.0	63.0	63.5
Education, %			
<HS	9.2	6.0	7.7
≥HS	90.8	94.0	92.3
OA, %			
Hip	12.0	5.6	8.7
Knee	70.0	66.7	68.3
Both	18.0	27.8	23.1
ACR functional class, %			
I	28.0	37.0	32.7
II	52.0	55.0	53.8
III	20.0	7.4	13.5
Duration of arthritis, years			
Mean	8.4	7.6	7.9
SD	9.7	5.9	7.9
Minimum	0.7	0.5	0.5
Median	4.9	6.1	6.0
Maximum	39.0	26.0	39.0

* No significant differences between treatment groups, $P > 0.05$. HS = high school; OA = osteoarthritis; ACR = American College of Rheumatology (formerly the American Rheumatism Association).

Table 2. Mean ± SD for Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) and Medical Outcomes Study 36-item Short Form Health Survey (SF-36) scales*

Scale	Baseline		Day 7		Day 14		Day 28	
	Placebo n = 50	Ibuprofen n = 54	Placebo n = 50	Ibuprofen n = 50	Placebo n = 47	Ibuprofen n = 47	Placebo n = 46	Ibuprofen n = 49
WOMAC scales								
Pain	64.6 ± 24.4	59.7 ± 21.8	66.0 ± 23.0	71.6 ± 23.1	69.8 ± 25.5	74.1 ± 22.3	70.3 ± 27.6	75.9 ± 23.0
Physical functioning	63.2 ± 24.2	59.2 ± 22.0	64.3 ± 27.2	69.4 ± 22.1	66.2 ± 27.4	71.8 ± 23.7	68.9 ± 28.1	72.8 ± 22.6
Stiffness	52.9 ± 30.3	46.1 ± 25.8	53.7 ± 29.9	59.0 ± 25.9	54.6 ± 29.7	57.1 ± 26.0	57.2 ± 30.3	67.6 ± 26.9
Total	61.7 ± 24.3	56.7 ± 23.9	62.8 ± 26.7	68.3 ± 21.9	65.3 ± 26.4	71.8 ± 22.2	67.0 ± 26.7	73.1 ± 22.4
SF-36 scales†								
Pain	53.8 ± 20.1	48.7 ± 16.4	54.4 ± 20.1	56.7 ± 21.1	55.7 ± 21.7	59.3 ± 23.5	55.3 ± 21.4	61.5 ± 23.7
Physical functioning	52.3 ± 23.4	50.3 ± 23.3	51.3 ± 24.6	51.9 ± 23.0	52.0 ± 24.4	56.2 ± 25.1	50.5 ± 24.7	52.1 ± 25.7
Role physical	63.5 ± 41.4	58.6 ± 42.3	61.5 ± 39.2	61.0 ± 41.7	61.2 ± 44.8	67.6 ± 38.5	58.2 ± 41.9	65.3 ± 42.3
Mental function	76.3 ± 16.7	79.7 ± 15.0	79.7 ± 13.0	78.2 ± 16.6	81.2 ± 15.7	79.5 ± 17.3	79.7 ± 16.4	80.9 ± 17.6
Role mental	74.7 ± 39.0	75.9 ± 38.5	84.0 ± 31.0	76.3 ± 33.6	81.6 ± 35.3	79.4 ± 36.1	81.9 ± 34.9	73.5 ± 39.1
Social function	80.5 ± 21.5	84.0 ± 23.1	83.5 ± 22.6	80.5 ± 24.5	82.4 ± 22.4	86.2 ± 21.2	85.3 ± 20.6	81.9 ± 25.0
Energy/vitality	51.8 ± 17.7	56.5 ± 21.1	55.5 ± 18.0	57.3 ± 20.4	54.3 ± 16.9	61.6 ± 22.4	50.8 ± 20.7	62.4 ± 22.5
General health	68.9 ± 17.9	66.3 ± 20.9	68.4 ± 16.4	66.1 ± 20.4	69.0 ± 17.6	65.9 ± 22.2	67.6 ± 16.5	62.4 ± 22.3
MCS	53.0 ± 9.9	55.6 ± 9.1	56.0 ± 8.4	54.7 ± 11.1	55.7 ± 9.8	55.4 ± 9.6	55.6 ± 8.4	55.0 ± 10.2
PCS	38.9 ± 13.3	36.6 ± 9.0	37.7 ± 10.5	36.3 ± 10.4	37.3 ± 10.1	39.9 ± 10.5	37.2 ± 10.3	39.0 ± 11.2
Global satisfaction	66.0 ± 29.1	57.6 ± 33.1	55.0 ± 27.4	72.2 ± 30.2	59.8 ± 28.9	69.0 ± 32.1	61.6 ± 29.1	73.7 ± 26.6

* No significant differences between treatment groups at baseline, $P > 0.05$.
† MCS = mental component summary; PCS = physical component summary.

groups were also strong for the two pain scales (ibuprofen r = 0.63 and placebo r = 0.65) and the two physical functioning scales (ibuprofen r = 0.71 and placebo r = 0.72). The WOMAC total score and the SF-36 physical component summary also were highly correlated (r = 0.75), and this correlation was consistent across treatment (ibuprofen r = 0.70, placebo r = 0.78). The WOMAC pain and physical function scales both showed strong correlation with the SF-36 physical component summary (r = 0.71–0.75) at day 14 as well. Scores at day 14 for the other components of the SF-36 showed low to moderate correlations (r = 0.20–0.50) with the WOMAC pain and physical function scales. Correlations of the day 14 WOMAC and the SF-36 scores with global satisfaction were moderate in magnitude ranging from 0.25 to 0.44 for day 14 scores.

Analyses to assess longitudinal congruent validity showed only moderate correlations between the WOMAC and SF-36 (Figure 2). The correlation coefficients between change scores at day 14 on the two pain scales and the two physical function scales were r = 0.26 (ibuprofen r = 0.23, placebo r = 0.21) and r = 0.25 (ibuprofen r = 0.24, placebo r = 0.18), respectively. Change in the WOMAC total score was most highly correlated with the SF-36 change scores for bodily pain, physical functioning, and the physical component summary score (r = 0.32–0.34). When these analyses were stratified by treatment, correlations for the ibuprofen group ranged from 0.13 to 0.50 compared with 0.12 to 0.40 for the placebo group. Correlations of change scores for the WOMAC with the other SF-36 domains ranged from 0.05 to 0.20. Change scores for global satisfaction were moderately correlated with the WOMAC and SF-36 change scores, with estimated coefficients ranging from 0.22 to 0.30.

Results for the ANOVA analyses are shown in Tables 3 and 4. In general, the WOMAC was more responsive (within-treatment differences, Table 3) and showed greater effect sizes (between-treatment differences, Table 4) than the corresponding scales on the SF-36. This was particularly true for the WOMAC physical function and total scores; within- and between-treatment differences were highly significant at days 7, 14, and 28 with these measures. The SF-36 bodily pain scale performed almost as well as the WOMAC pain scale. With respect to the other SF-36 scales, only the energy/vitality scale (Table 2) showed some improvement from baseline with ibuprofen, and this improvement appeared to be better than placebo. The improvement in energy/vitality for the ibuprofen group did not appear until day 14, whereas improvements in both pain scales appeared within the first 7 days. Neither questionnaire detected significant changes in the placebo group, as expected (Table 3).

Patients receiving ibuprofen had consistently greater global satisfaction with medication across the study period than placebo patients; however, the magnitude of this difference was not consistent across time. Patients in the ibuprofen group showed

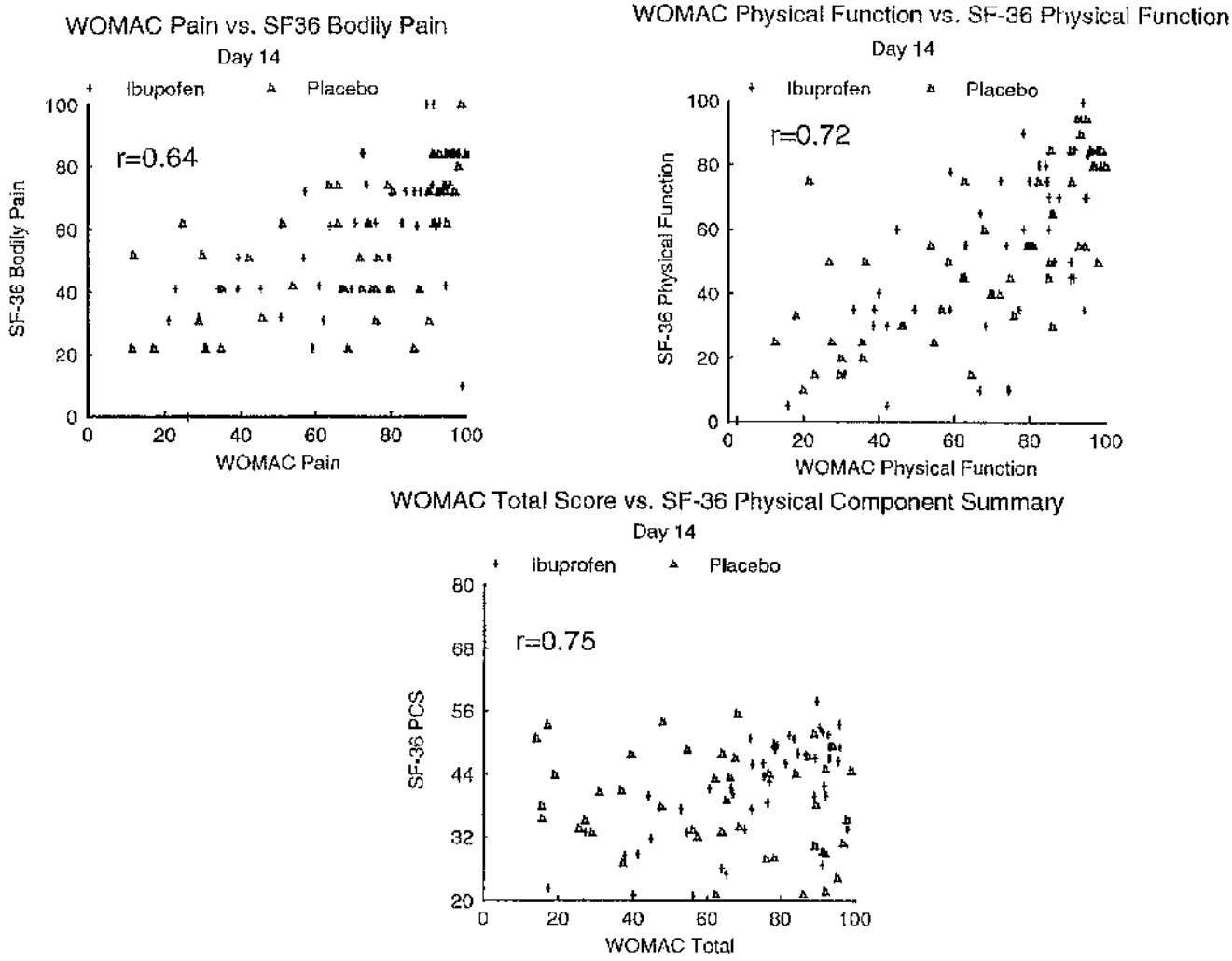


Figure 1. Analyses to assess congruent validity. WOMAC = Western Ontario and McMaster Universities Osteoarthritis Index; SF-36 = Medical Outcomes Study 36-item Short Form Health Survey; PCS = physical component summary. r = Spearman rank correlation coefficient.

increased satisfaction with treatment by day 7. The placebo group showed reduced satisfaction with treatment compared with baseline over the length of the study period.

DISCUSSION

Previous evaluations of the measurement characteristics of the WOMAC and the SF-36 in OA patients have been done following joint replacement surgery (8,9,17,18), an intervention that results in a dramatic change in disease status. In that setting, the WOMAC performed better at discriminating between patients with varying levels of knee problems, and the SF-36 performed better at discriminating between patients with varying levels of general health status and comorbidities.

In clinical trials of pharmaceutical interventions for OA, changes in disease status following therapy may be less dramatic than those seen following joint replacement surgery. Therefore, it is important to choose the most sensitive instrument available to most efficiently detect differences among treatments. To our knowledge, this is the first published report to evaluate the responsiveness and relative effect size of the WOMAC and applicable scales from the SF-36 in a randomized, clinical trial of a pharmaceutical intervention for OA.

In this study, the pain, physical function, and

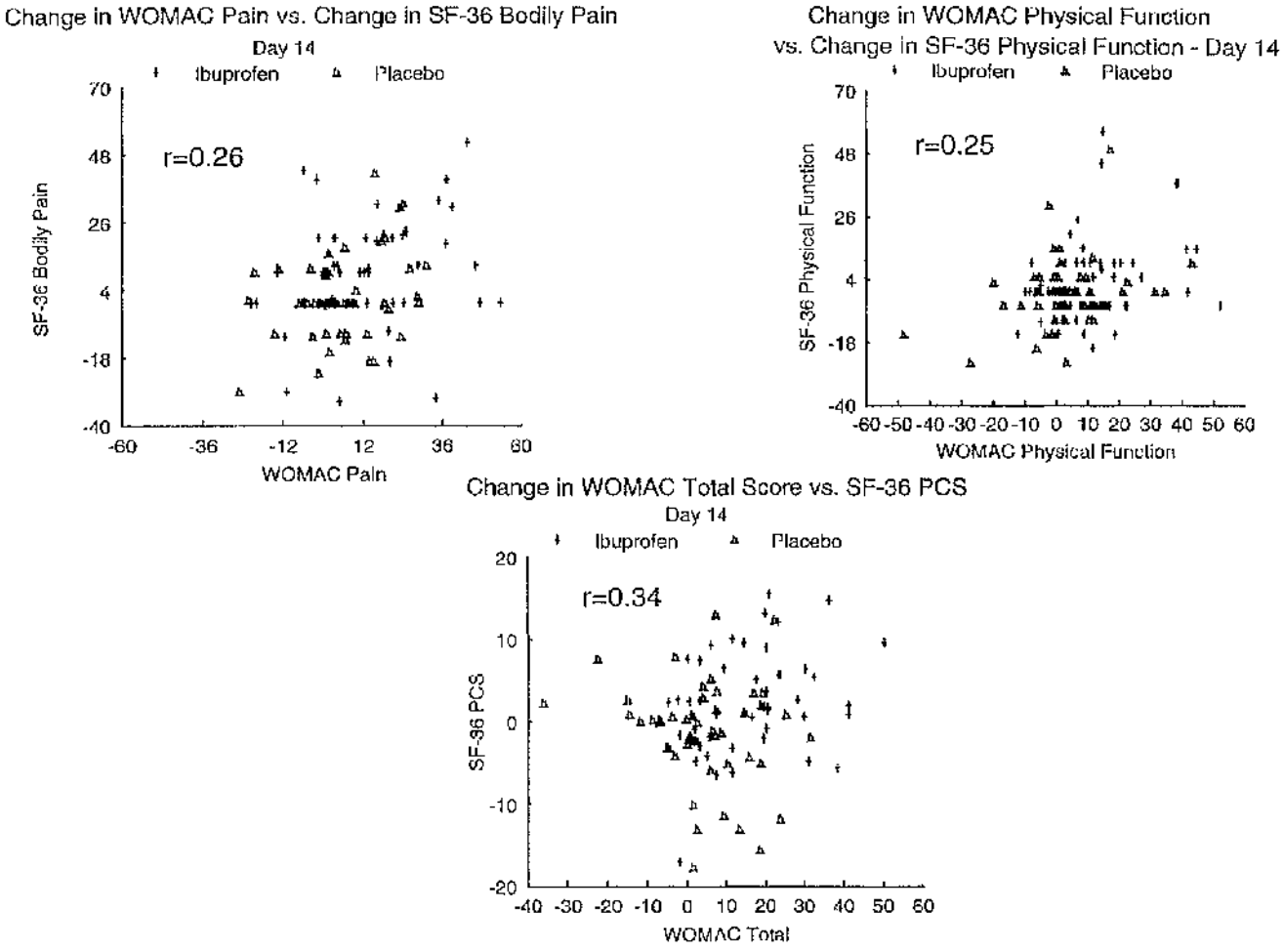


Figure 2. Analyses to assess longitudinal congruent validity. WOMAC = Western Ontario and McMaster Universities Osteoarthritis Index; SF-36 = Medical Outcomes Study 36-item Short Form Health Survey; PCS = physical component summary. r = Spearman rank correlation coefficient.

total score from the WOMAC and the bodily pain scale from the SF-36 were able to detect response to therapy with ibuprofen and show differences between active and placebo treatment in patients with OA of the hip and/or knee. However, the WOMAC proved to be the more efficient of the two instruments. This was demonstrated by the more rapid and greater sustained increase in scores following initiation of therapy, and by the generally larger effect sizes for the treatment group differences with the WOMAC pain, physical function, and total scores compared with those of the SF-36 bodily pain, physical function, and physical component summary scores, respectively. Because this comparison was based on a 28-day study, these results may not pertain to trials of longer duration.

Features of the study design may have posed problems in the use of the SF-36. Some patients were receiving therapy for OA prior to entering the trial, which was discontinued during the 4-day run-in period. Baseline (day 0) assessments were made at the end of the placebo run-in period. Therefore, in responding to questions on the SF-36 (acute version), which has a recall period of one week, their responses necessarily reflected some days during which their prior therapy may still have been exerting some effect. This would not have been a problem with the WOMAC, which has a recall period of 48 hours.

The recall period for this version of the WOMAC was 48 hours compared with the 1-week version of the SF-36. Studies of the recall period for the WOMAC suggest that there is no time dependency in

Table 3. Evaluation of responsiveness of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) and corresponding scales and summary scales on the Medical Outcomes Study 36-item Short Form Health Survey (SF-36)*

Treatment day	Treatment group	Change score statistic	WOMAC pain	SF-36 BP	WOMAC PF	SF-36 PF	WOMAC total	SF-36 PCS	Global satisfaction
Day 7	Ibuprofen	Mean	11.6†	8.1†	10.4†	0.7	11.9†	1.5	14.5†
	(n = 50)	SD	17.3	17.3	14.2	13.3	14.0	6.6	31.4
	Placebo	Mean	1.4	0.8	1.0	−1.0	1.1	2.0	−11.0†
	(n = 50)	SD	17.9	12.9	12.0	12.4	11.9	9.0	24.1
Day 14	Ibuprofen	Mean	13.1†	9.6†	12.0†	3.8	14.3†	2.4	10.1
	(n = 47)	SD	17.3	17.9	14.2	14.9	13.4	6.3	33.6
	Placebo	Mean	5.9	1.9	3.3	−0.4	4.2	−1.4	−5.9
	(n = 47)	SD	13.2	14.0	15.2	12.2	12.6	6.4	34.5
Day 28	Ibuprofen	Mean	15.4†	11.9†	13.6†	1.4	16.0†	2.0	14.0†
	(n = 49)	SD	16.5	20.5	15.2	15.7	14.1	7.3	29.8
	Placebo	Mean	5.3	1.0	5.2†	−3.7	5.5	−0.4	−5.2
	(n = 46)	SD	18.2	14.1	15.7	11.9	15.6	11.7	32.4

* BP = bodily pain; PF = physical function; PCS = physical component summary.
† Test of mean change ≠ 0, $P \leq 0.05$.

the WOMAC for periods between 2 days and 2 weeks (12). This previous research suggests that the difference in the recall periods should not be an issue in comparisons of these two instruments.

Two component summary measures for the SF-36 were also computed in this study. The scoring algorithm for the physical component summary and mental component summary are based on estimated factor coefficients from a survey of the general US population (14). The use of the summary scores to assess responsiveness in our study may not be valid since the study population is not representative of the general US population. However, this bias should not affect the results for the comparisons between ibuprofen and placebo.

As expected, the global satisfaction with medication question showed ibuprofen patients were more satisfied with treatment than placebo patients; however, the estimates of variance were large relative to their mean. This result is not surprising, because satisfaction encompasses physical and emotional constructs as well as effects from adverse events. Patients experiencing medication side effects may have lower satisfaction even if their pain and physical functioning improved substantially. This could cause increased variability in this measure.

There are several possible reasons why the SF-36 was not as responsive as the WOMAC in this study. The VAS response format of the WOMAC may account for the greater responsiveness of this instrument. A Likert scaling version of the WOMAC does exist and has been shown to be slightly less sensitive than the VAS version (12). It would be difficult to say whether the Likert version of the WOMAC physical function scale would have been as responsive in this setting. In addition, patients with an ACR functional

Table 4. Results of relative effect size of ibuprofen versus placebo [pooled SD]*

Change score	WOMAC pain	SF-36 BP	WOMAC PF	SF-36 PF	WOMAC total	SF-36 PCS	Global satisfaction
Day 7	9.4†	6.3†	8.7†	1.1	10.0†	2.4	26.3†
	17.4	15.5	13.3	13.1	13.2	6.0	27.5
	0.54	0.41	0.65	0.08	0.76	0.40	0.96
Day 14	8.6	5.8	8.4†	3.6	9.6†	3.2†	13.9
	15.9	15.6	14.8	13.8	13.0	6.5	33.7
	0.42	0.37	0.57	0.26	0.74	0.49	0.41
Day 28	6.9†	10.1†	7.9†	4.5	9.5†	4.0†	19.4†
	17.7	18.3	15.9	14.5	14.9	7.1	32.1
	0.50	0.55	0.50	0.28	0.64	0.56	0.60

* WOMAC = Western Ontario and McMaster Universities Osteoarthritis Index; SF-36 = Medical Outcomes Study 36-item Short Form Health Survey; BP = bodily pain; PF = physical function; PCS = physical component summary.
† Test of treatment difference ≠ 0, $P \leq 0.05$.

class of IV were not included in the study; the SF-36 physical function or physical component summary may be just as sensitive as the WOMAC in this type of patient.

The SF-36 is a generic health status instrument that assesses psychosocial constructs as well as physical constructs. Conceptually, improvements in physical constructs could result in future improvements in psychosocial and mental health components. In this study, no changes were observed in the remaining scales of the SF-36 except for the energy and vitality scale. It is possible that the duration of this study was too brief to detect any improvement in the scales of the SF-36 that reflect mental and psychosocial constructs.

Both the WOMAC and the SF-36 have similar constructs of pain and physical functioning. The WOMAC pain subscale was specifically developed for pain related to OA. However, the SF-36 pain question refers to general bodily pain. If ibuprofen was more effective for OA pain than for pain from other sources, this could explain the differences in responsiveness between the instruments in this study. The items for the SF-36 and WOMAC physical function scales are very similar. The lack of response on the SF-36 physical function scale in this study could be explained by the lack of sufficiently sensitive response categories for these OA patients.

From this evaluation, it appears that the WOMAC is highly responsive to changes in OA disease status and is more efficient than the corresponding scales of the SF-36 in short-term clinical trials of pharmaceutical therapy for OA. As a result, use of the WOMAC to demonstrate efficacy in clinical drug trials of OA would allow smaller sample sizes and resource savings.

REFERENCES

1. Meenan RF, Gertman PM, Mason JH. Measuring health status in arthritis: the Arthritis Impact Measurement Scales. Arthritis Rheum 1980;23:146–52.
2. Fries JF, Spitz PW, Young DY. The dimensions of health outcomes: the Health Assessment Questionnaire, disability and pain scales. J Rheumatol 1982;9: 789–93.
3. Doyle DV, Dieppe PA, Scott J, Huskisson EC. An articular index for the assessment of osteoarthritis. Ann Rheum Dis 1981;40:75–8.
4. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt W. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988;15:1833–40.
5. Lequesne M. European guidelines for clinical trials of new antirheumatic drugs. EULAR Bull 1980;9 Suppl 6:171–5.
6. Brooks RH, Callahan LF, Pincus T. Use of self-report activities of daily living questionnaires in osteoarthritis. Arthritis Care Res 1988;1:23–32.
7. Callahan LF, Smith WJ, Pincus T. Self-report questionnaires in five rheumatic diseases: comparisons of health status constructs and associations with formal educational level. Arthritis Care Res 1989;2:122–31.
8. Hawker G, Melfi C, Paul J, Green R, Bombardier C. Comparison of a generic (SF-36) and a disease specific (WOMAC) instrument in the measurement of outcomes after knee replacement surgery. J Rheumatol 1995;22:1193–6.
9. Bombardier C, Melfi C, Paul J, Green R, Hawker G, Coyte P. Comparison of a generic and a disease-specific measure of pain and physical function after knee replacement surgery. Med Care 1995;33 Suppl 4:AS131–44.
10. Altman R, Alarcón C, Appelrouth D, Bloch D, Borenstein D, Brandt K, et al. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. Arthritis Rheum 1991; 34:505–14.
11. Altman R, Asch E, Bloch D, Bole C, Borenstein D, Brandt K, et al. Development of criteria for the classification and reporting of osteoarthritis: classification of osteoarthritis of the knee. Arthritis Rheum 1986;29: 1039–49.
12. Bellamy N. WOMAC Osteoarthritis Index—a user's guide. London (Canada): Victoria Hospital; 1995.
13. Ware JE, Snow KK, Kosinski M, Gandek B. SF-36 health survey: manual and interpretation guide. Boston: The Health Institute; 1993.
14. Ware JE, Kosinski M, Keller SD. SF-36 physical and mental health summary scales: a user's manual. Boston: The Health Institute; 1994.
15. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. Oxford: Oxford University Press; 1995.
16. SAS Institute. SAS/STAT user's guide, version 6. Vol. 2. 4th ed. Cary (NC): SAS Institute; 1989.
17. Brooks PM, Bachmeier CJM, March LM, Cross M, Courtenay B, Schwarz J, et al. A comparison of WOMAC and MOS SF-36 in OA patients undergoing joint replacement [abstract]. Arthritis Rheum 1997;40 Suppl 9:S110.
18. Saugha O, Theiler R, Schären S, Stucki G. Comparative responsiveness of health status instruments and clinical variables in osteoarthritis [abstract]. Arthritis Rheum 1997;40 Suppl 9:S235.

SECTION 5 - Applications in Clinical Research

The value of an outcome measure is reflected in its performance in post-validation applications. The WOMAC Index has been widely used since its development in 1982. The studies presented in this section of the thesis are interventional studies in which the originator of the WOMAC Index was either the principal investigator or a co-investigator, and in which the WOMAC Index was used as a primary or secondary outcome measure (12-18). Some of the studies were also used to test measurement hypotheses or evaluate other aspects of the WOMAC Index.

The RCT of sodium meclofenamate versus diclofenac sodium was the first post-validation application of the WOMAC VA3.0 Index, and used a joint-targeted version of the WOMAC Index (12). Statistically significant within-group improvements were detected in both treatment groups, and between-group differences favouring meclomen were detected in the WOMAC pain and stiffness subscales. In addition to detecting between-group differences between two active agents, the results of this study also suggested that the WOMAC Index was similar in responsiveness, as assessed by the relative efficiency statistic, to the Doyle Index and the Lequesne Index of Clinical Severity (12). Finally the study provided an improved estimate of the standard deviations of WOMAC subscale scores for calculating sample size for future studies using the WOMAC Index (12).

The RCT of tenoxicam versus diclofenac was the first Canadian study to compare the two agents in knee OA (13). The importance of this study is that separate WOMAC questionnaires were completed for each knee, an innovation that informed the development of the WOMAC 3.1 series of questionnaires which are joint-specific, and require the patient to rate their symptom experience in a single joint. Statistically significant within-group improvements were detected in both treatment groups on all three WOMAC subscales. No significant between-group differences were observed between these two active agents (13).

The RCT of controlled-release codeine versus placebo permitted observation of the performance of the WOMAC Index in a different research environment, that of a complex analgesic (14). Statistically significant between-group differences in favour of controlled-release codeine were detected by all three WOMAC subscales (14). This study suggested that therapeutic agents having a primarily analgesic effect might produce clinical benefit not only on pain, but also on stiffness and physical function, and be detectable by the relevant subscales of the WOMAC Index.

The RCT of diacerein versus placebo permitted observation of the performance of the WOMAC Index in a clinical study involving a symptom modifying slow-acting drug for OA (SADOA)(15). In this trial the WOMAC Index was used as a secondary outcome measure, and detected statistically significant differences versus placebo and in favour of diacerein 100 mg/day in the intent-to-treat (ITT) analysis on all three WOMAC subscales (15).

An open-label study of an education-driven skills acquisition programme provided opportunity to study the performance of the pain and function subscales of the WOMAC Index in yet another research environment, involving the treatment of knee OA patients, in general practice, with a viscosupplementation product, hylan G-F 20 (16). Statistically significant within-patient improvements were noted on both WOMAC

subscales (16). In addition, this study validated a patient global assessment question, that subsequently became part of two other indices, the Western Ontario Measurement Battery (WOMBAT) (16), and the Osteoarthritis Global Index (OGI).

The pragmatic RCT of the effectiveness, cost-effectiveness and cost-utility of "Appropriate Care + hylan G-F 20 versus Appropriate Care without hylan G-F 20" provided an opportunity to employ the WOMAC Index to pre-specify a definition of a responder, in order to perform a comparative analysis based on responder criteria (17), and also to employ the WOMAC Index in a cost-effectiveness evaluation (18). The time frame for the WOMAC Index for this study was set at one-month. The WOMAC was able to detect a clinically important, statistically significant between-group difference in favour of "Appropriate Care + hylan G-F 20" (17). In the cost-effectiveness analysis, the WOMAC Index was used to specify a definition for an improved patient, on the basis of which, the incremental cost of a patient improved over one year was calculated to be $2505 CDN (societal) and $9930 CDN (health care system)(18).

Collectively these post-validation applications of the WOMAC Index confirmed the responsiveness of the three subscales of the WOMAC Index in such diverse research environments as trials of NSAIDs, a SADOA class agent, a complex analgesic and viscosuppplementation with hylan G-F 20. These studies also suggested that the WOMAC was a relatively robust measure, since its performance was maintained when focussing the patient's symptom severity rating on only a single joint (12-18), and extending the time frame for recall to one month (17,18). These studies also confirmed the feasibility of applying the WOMAC Index in different clinical practice environments including orthopaedics (17,18), rheumatology (12-15, 17,18) and general practice (16).

References:

12. Bellamy, N., Kean, W.F., Buchanan, W.W., Gerecz-Simon, E. and Campbell, J. Double blind randomized controlled trial of sodium meclofenamate (meclomen) and diclofenac sodium (voltaren): Post validation reapplication of the WOMAC Osteoarthritis Index. Journal of Rheumatology 1992;19:153-159.

13. Bellamy, N., Buchanan, W.W., Chalmers, A., Ford, P.M., Kean, W.F., Kraag, G.K., Gerecz-Simon, E. and Campbell, J. A multicentre study of tenoxicam and diclofenac in patients with osteoarthritis of the knee. Journal of Rheumatology 1993;20:999-1004.

14. Peloso, P.M., Bellamy, N., Bensen, W., Thomson, G.D., Harsanyi, Z., Babul, N., and Darke, A.C. Double blind randomized placebo control trial of controlled-release codeine in the treatment of osteoarthritis of the hip or knee. Journal of Rheumatology 2000;27(3):764-771.

15. Pelletier, J-P., Yaron, M., Haraoui, B., Cohen, P., Nahir, M.A., Choquette, D., Wigler, I., Rosner, I.A., Beaulieu, A.D., and The Diacerein Study Group. Efficacy and safety of diacerein in osteoarthritis of the knee. Arthritis & Rheumatism 2000;43(10):2339-2348.

16. Bellamy, N., Goldstein. L., and Tekanoff, R. CME-Driven skills acquisition and impact on improved patient outcomes. Journal of Continuing Education in the Health Professions 2000;20:52-61.

17. Raynauld J.P., Torrance G.W., Walker V., Goldsmith C.H., Bellamy N., Band P.A., Schultz M. and Tugwell P. A prospective, randomized, health outcomes trial evaluating the incorporation of hylan G-F 20 into the treatment paradigm for patients with knee osteoarthritis (part 1 of 2): clinical results. Osteoarthritis and Cartilage 2002;10(7):506-517.

18. Torrance G.W., Raynauld J.P., Walker V., Goldsmith C.H., Bellamy N., Band P.A., Schultz M. and Tugwell P. A prospective, randomized, health outcomes trial evaluating the incorporation of hylan G-F 20 into the treatment paradigm for patients with knee osteoarthritis (part 2 of 2): economic results. Osteoarthritis and Cartilage 2002;10(7):518-527.

Double Blind Randomized Controlled Trial of Sodium Meclofenamate (Meclomen) and Diclofenac Sodium (Voltaren): Post Validation Reapplication of the WOMAC Osteoarthritis Index

NICHOLAS BELLAMY, WALTER F. KEAN, W. WATSON BUCHANAN, ERIKA GERECZ-SIMON, and JANE CAMPBELL

Abstract. **Following several years of development and validation, we applied the WOMAC Osteoarthritis Index as the principal outcome measure in a double blind randomized parallel trial of Meclomen (100 mg po tid) and Voltaren (25 mg po tid). Statistically significant improvements in clinical status were noted in both treatment groups. At the doses studied, between drug differences favoring Meclomen were observed in pain and stiffness, no difference being noted in physical function. No significant between drug difference was noted in tolerability at these same doses. Our study also demonstrated that the relative efficiency of WOMAC was similar to that of the Lequesne and Doyle indices. Finally, we defined the standard deviation necessary to calculate sample size for future studies using the WOMAC index, both for studies based on static scores and those based on change scores.** (*J Rheumatol 1992;*19:153–9)

Key Indexing Terms:
CLINICAL METROLOGY OSTEOARTHRITIS NSAID

In recent years we identified significant variability in the outcome measures used in osteoarthritis (OA) clinical trials[1]. Furthermore, we have noted that the ability to detect differences between an active drug and a placebo is greater than that between 2 active drugs. Nevertheless, some traditional measures (e.g., 50′ walk time) cannot predictably detect between treatment differences even in placebo controlled studies. In an attempt to standardize the measurement of pain, stiffness, and physical function in antirheumatic drug studies in OA, and to develop an instrument of superior responsiveness that might be capable of detecting between drug differences, we developed a tridimensional self-administered questionnaire probing patient relevant clinically important outcomes termed the Western Ontario and McMaster Universities (WOMAC) Osteoarthritis Index[2-7]. The instrument consists of 24 questions, grouped into 3 (3.0) subscales (pain, stiffness and physical function). There are 4 versions of this index (LK3.0, LK3.0S, VA 3.0, VA 3.0S), which differ only in the type of scale on which the response is scored (LK = 5 point Likert scale, VA = 10 cm horizontal visual analog scale) and whether the patient is asked to identify one signal (S) item in each subscale to permit a signal rather than an aggregate approach to measurement[7]. Our major interest currently is to assess further other clinimetric properties of the WOMAC Index. We have reapplied WOMAC-VA 3.0, therefore, in a comparative study of Meclomen versus Voltaren in patients with primary OA of the knee. The efficacy and tolerability of Voltaren have been described[8]. Meclomen is not currently available in Canada and was designated, therefore, the "test" drug. However, studies performed elsewhere attest to the efficacy of Meclomen in the treatment of OA of the knee[9-11]. Our goals were 4-fold: (1) to compare the efficacy of Meclomen 100 mg per os (po) 3 times daily (tid) and Voltaren 25 mg po tid using WOMAC as the principal outcome measure, (2) to compare the 2 drugs with respect to any clinical or laboratory toxicity, (3) to examine the relative efficiency of WOMAC compared to 2 other indices in assessing the clinical response, (4) to generate mean and standard deviation estimates for static scores and change scores for WOMAC-VA 3.0 in order to calculate sample size for future antirheumatic drug studies.

MATERIALS AND METHODS

Sixty-nine consecutive consenting outpatients with primary OA of the knee were entered into the study. The following inclusion and exclusion criteria were applied: Inclusion — symptomatic primary OA of at least one knee warranting treatment with nonsteroidal antiinflammatory drugs (NSAID),

From the Department of Medicine, The University of Western Ontario, London, and McMaster University, Hamilton, ON, Canada.

Supported in part by a grant from Parke Davis, Division of Warner-Lambert Canada Inc

N. Bellamy, MD, MSc, FRCP(Glas,Edin), FRCPC, FACP, Associate Professor of Medicine, Epidemiology and Biostatistics, University of Western Ontario; W.F. Kean, MD, FRCPC, FACP, FARA, Associate Clinical Professor; W.W. Buchanan, MD, FRCP(Glas,Edin), FRCPC, FACP, Professor of Medicine; E. Gerecz-Simon, MD, Codirector, Musculoskeletal Pain Research, Service of Rheumatology, McMaster University; J. Campbell, BA, Research Assistant, University of Western Ontario.

Address reprint requests to Dr. N. Bellamy, Division of Rheumatology, Suite 402A, Victoria Hospital, Westminster Tower, 800 Commissioners Rd. E., London, ON N6A 4G5 Canada.

Submitted December 3, 1990 revision accepted September 9, 1991.

age 40–65 years, radiographic evidence of at least one of the following: osteophytes, joint space narrowing, periarticular sclerosis, subchondral cysts. A symptomatic patient was defined as a patient with pain and disability from OA of the knee. Primary OA was defined by the presence of typical joint symptoms (pain, stiffness, disability), signs (bony crepitus), and radiographic findings of OA as illustrated in the *Standard Atlas of Radiographs*[12]. Patients with an atypical distribution of OA were excluded if they had evidence of chondrocalcinosis or elevated ferritin, thyroid stimulating hormone or calcium level. All radiographs were graded according to *Standard Atlas* criteria (Grades 1–4). Patients were not enrolled if their disease was confined only to the patellofemoral compartment. Exclusion — any other form of joint disease or prior knee replacement surgery, pregnancy or lactation, active peptic ulceration in the past 2 years or upper gastrointestinal bleeding in the past 5 years, severe uncontrolled cardiorespiratory insufficiency, blood urea nitrogen (BUN) > 30 mg/dl, aspartate aminotransferase (AST) > 50 units/ml, documented allergy to aspirin or any other NSAID, use of oral anticoagulants, systemic or intraarticular corticosteroid use in the past 3 months.

Following initial screening (medical history and clinical exmaination), there was a 3 to 7 day NSAID-free washout period during which patients were instructed to cease all NSAID therapy, to take placebo capsules tid and were permitted to take monitored quantities of acetaminophen. Patients were then reassessed (baseline) and thereafter randomized to receive either Meclomen 100 mg po tid or Voltaren 25 mg po tid for the next 6 weeks. Since the conventional dose range of Meclomen is 200–400 mg/day, and for Voltaren 50–150 mg/day, this study represents a comparison of a midrange dose of Meclomen *against* a relatively low dose of Voltaren.

The allocation schedule used a blocked randomization technique (blocking factor = 10). Both assessors and patients were blind to the allocation until after completion of the study. All study capsules (Meclomen, Voltaren and placebo) were identical in appearance. During the active treatment phase, patients were permitted to take supplementary analgesia sparingly (if required, acetaminophen up to 6 × 325 mg/day). Compliance to study medication and analgesic ingestion were verified by pill counting. Patients were assessed at 5 points during the trial: screening, 0 (baseline), 2, 4, and 6 (termination) weeks. The following clinical assessments were made at each time point: (1) WOMAC-VA3.0 OA Index[2]; (2) Lequesne Knee Index[13] (a self-administered 11 item questionnaire probing pain, stiffness, maximum distance walked and activities of daily living); (3) Doyle Articular Index[14] (a modification of the Ritchie Index in which 48 joints or joint units are assessed by palpation for tenderness on a scale of 0 = no tenderness, 1+ patient complains of pain, 2+ patient complains of pain and winces, 3+ patient complains of pain, winces and withdraws); (4) Night pain (4 point Likert scale); (5) Pain on walking (4 point); (6) Starting pain (4 point); (7) Stress pain (4 point); (8) Swelling (4 point); (9) Degree of tenderness (4 point); (10) Limitation of movement (4 point); (11) Maximum knee flexion (degrees); (12) Maximum knee extension (degrees); (13) Overall assessment (separate patient and physician assessments) (5 point); (14) Comparative assessment of condition to previous visit (separate patient and physician global assessments) (3 point). At the screening visit, one knee was identified as the study joint (i.e., the worst knee as the primary focus of measurement for future assessment in the study). Records were kept of concomitant medications, study medication, blood pressure, body weight and temperature. Adverse reactions were ascertained by open ended indirect questioning. Routine hematology, clinical chemistry and urinalysis were completed at screening, Week 2 and termination.

After study completion, double data entry was performed in collaboration with Innovus Inc., Hamilton, ON, using the SIR (Scientific Information Retrieval) database[15] and data analysis conducted using SAS Version 6.03[16] and BMDP-UX software packages[17]. Data were checked for validity, normality, skewness and kurtosis. Continuous variables were analyzed by analysis of covariance using the respective baseline measures as the covariates. Descriptive analyses were conducted on all variables defining means and standard deviations. Categorical outcome measures were analyzed using χ^2 tests to determine differences between the treatment groups, and McNemar's χ^2 test to determine differences between treatment visits. The Bartlett test was used to test for homogeneity of variance between centers and treatments at baseline. In all statistical tests, the level of Type I error (2-tailed) was set at 0.05. Although 16 separate analyses were made we have made no correction for multiple comparisons. This is based on the following considerations: (1) A single primary efficacy variable was defined *a priori*. (2) Efficacy variables were highly correlated. (3) The extent of any correction remains contentious both statistically and clinically. (4) Precedent indicates that in the OA literature such correction is rarely made[8], and (5) Correction for multiple comparisons is generally not appropriate in exploratory studies[18]. A decision had been made *a priori* to use the WOMAC OA Index (pain subscale) as the principal outcome measure, all other variables being relegated to secondary outcome status. This allowed the relative efficacy of the 2 active drugs to be defined by WOMAC, but also permitted an examination of the Relative Efficiency (RE)[19] of WOMAC versus the Doyle and Lequesne indices. RE is defined as the square of the ratio of 2 t values, i.e., RE (WOMAC vs Doyle) = (t WOMAC/t DOYLE)^2. If the RE is > 1.0, the instrument in the numerator can be inferred to be the more responsive measure of outcome requiring smaller sample sizes and/or detecting smaller effect sizes than the instrument in the denominator.

RESULTS

Sixty-six patients entered the trial of whom 5 were excluded from analysis for the following reasons: intolerance = 1 (placebo), noncompliance = 4. Of the remaining 61 patients, 30 received Meclomen and 31 received Voltaren. The disease and demographic profiles of participating patients are illustrated in Table 1. No significant between group differences were noted in age, disease duration, sex, or in any of the disease characteristics. The baseline and termination status (mean and standard deviation) of the outcome variables is illustrated in Table 2 (indices) and Table 3 (study joint and global/overall measures). Statistically significant between group differences at baseline were noted in the pain ($p = 0.0313$), stiffness ($p = 0.0502$) and physical function ($p = 0.0080$) subscales of the WOMAC Index.

As noted, to account for these differences, data were analyzed using analysis of covariance (ANCOVA) techniques using the baseline values as the covariate.

Efficacy. Within group comparisons (baseline versus endpoint) were made for all outcome variables (Tables 2 and 3). Statistically significant improvements with treatment were noted as follows: Both drugs — WOMAC (all 3 subscales), Lequesne Index (all components), Doyle Index, walking pain, starting pain, stress pain, physician overall assessment and physician comparative assessment; Meclomen only — night pain and patient global assessment; Voltaren only — swelling. Although improvements occurred on other study joint measures (Table 3), none of these was statistically significant.

Between treatment comparisons were made for all outcome variables. A statistically significant difference favoring Meclomen was detected for the following variables: WOMAC Index (pain and stiffness), Lequesne Index (pain/discomfort and activities of daily living), Night Pain and Pain on Walking (Tables 2 and 3). The remaining index subscales, study joint measures and other overall/global

Table 1. *Postrandomization, preintervention comparison of treatment groups (mean ± SD)*

Variable	Meclomen (n = 30)	Diclofenac (n = 31)	p Value
Age (years)	55.73 (6.32)	56.97 (6.11)	NS
Disease duration (years)	6.04 (5.83)	8.03 (7.41)	NS
Sex (F, M)	19F, 11M	24F, 7M	NS
Bilateral (B) vs Unilateral (U) disease	24B, 6U	26B, 5U	NS
Most severely affected knee (L, R)	15R, 14L	16R, 15L	NS
Bony osteophytes	25	22	NS
Loss of joint space	23	28	NS
Eburnation of juxtaarticular bone	2	6	NS
Subchondral bony cysts	0	0	NS
Radiographic grading*			
Grade 1	5	3	NS
Grade 2	11	17	NS
Grade 3	13	8	NS
Grade 4	1	3	NS
Tibiofemoral disease only	12	9	NS
Patellofemoral disease only	0	0	NS
Tibiofemoral and patellofemoral disease	18	22	NS

* Atlas of Standard Radiographs (1963).

Table 2. *Mean values and standard deviations for OA indices*

Index		Baseline	Baseline vs Termination	Termination	Between Drugs at Termination
WOMAC					
Pain	(M)*	199.7 (105.3)	0.00	115.1 (117.6)	0.04
	(D)†	275.6 (97.6)	0.01	224.3 (140.0)	
Stiffness	(M)	94.4 (54.1)	0.00	53.9 (53.0)	0.02
	(D)	123.0 (37.8)	0.01	100.2 (57.2)	
Physical function	(M)	693.0 (336.8)	0.00	417.3 (392.3)	0.44
	(D)	952.8 (317.7)	0.00	768.6 (451.4)	
Lequesne					
Pain & discomfort	(M)	5.8 (1.2)	0.00	3.7 (2.3)	0.03
	(D)	6.2 (1.1)	0.00	5.2 (1.6)	
Walking	(M)	2.9 (1.7)	0.00	2.1 (1.7)	0.90
	(D)	3.0 (1.4)	0.00	2.5 (1.4)	
Activities of daily living	(M)	4.8 (1.5)	0.02	3.0 (2.0)	0.04
	(D)	5.2 (1.1)	0.00	4.5 (1.7)	
Total	(M)	13.5 (3.6)	0.00	8.8 (5.5)	0.06
	(D)	14.4 (2.9)	0.00	12.2 (4.1)	
Doyle	(M)	9.5 (9.1)	0.00	6.1 (9.7)	0.76
	(D)	8.9 (6.1)	0.00	6.2 (4.8)	

* M = Meclomen.
† D = Diclofenac.

measures failed to detect any significant difference between the 2 drugs. It should be noted that there were no differences in analgesic consumption to account for the between drug differences in perceived pain. Only one patient was withdrawn from each treatment group due to inefficacy (Table 4).

Tolerability. The frequency and nature of adverse reactions reported during the study are illustrated in Table 4. The majority of clinical adverse reactions reported were gastrointestinal in nature and accounted for about ¾ of all events in both treatment groups. Several patients reported more than one adverse event, resulting in numerically more adverse reactions on Meclomen (n=20) than Voltaren (n=15). The actual number of patients experiencing adverse reactions was small (Meclomen=8, Voltaren=7), and was not significantly different between the 2 treatment groups ($\chi^2 = 0.01$, $p = 0.9063$). No significant difference was detected in the number of patients withdrawn due to adverse reactions (Meclomen = 5, Voltaren = 2, $\chi^2 = 1.57$, $p = 0.2108$), or the number of patients with severe reactions (Meclomen = 2, Voltaren = 1, $\chi^2 = 0.12$, $p = 0.7274$). No significant

Table 3. *Mean values and standard deviations for unidimensional variables*

Outcome Variable		Baseline	Baseline vs Termination	Termination	Between Drugs at Termination
Night pain	(M)*	2.6 (0.8)	0.00	1.7 (1.0)	0.03
	(D)†	2.7 (0.9)	NS	2.2 (0.9)	
Walking pain	(M)	3.1 (0.6)	0.00	2.0 (0.9)	0.05
	(D)	3.1 (0.7)	0.03	2.5 (0.9)	
Starting pain	(M)	3.0 (0.9)	0.00	1.9 (0.9)	NS
	(D)	3.3 (0.8)	0.01	2.4 (1.0)	
Stress pain	(M)	2.4 (1.0)	0.04	1.6 (0.8)	NS
	(D)	2.5 (0.8)	0.01	1.7 (0.8)	
Swelling	(M)	1.4 (0.7)	NS	1.1 (0.3)	NS
	(D)	1.6 (0.8)	0.01	1.3 (0.6)	
Limitation of movement	(M)	1.2 (0.5)	NS	1.0 (0.2)	NS
	(D)	1.2 (0.4)	NS	1.1 (0.3)	
Knee flexion	(M)	113.0 (15.7)	NS	120.8 (9.2)	NS
	(D)	110.8 (18.0)	NS	117.0 (15.8)	
Knee extension	(M)	0.8 (2.5)	NS	0.2 (0.9)	NS
	(D)	0.8 (3.6)	NS	0.7 (2.1)	
Overall assessment	(M)	3.6 (0.8)	0.05	2.4 (0.9)	NS
(MD)	(D)	4.1 (0.8)	0.05	2.7 (1.1)	
Overall assessment (Pt)	(M)	3.7 (0.8)	0.00	2.4 (0.9)	NS
	(D)	4.9 (0.9)	0.01	2.7 (1.0)	
Comparative assessment	(M)	2.4 (0.5)	0.00	1.7 (0.7)	NS
(Global-MD)	(D)	2.6 (0.5)	0.01	1.9 (0.6)	
Comparative assessment	(M)	2.3 (0.6)	0.01	1.7 (0.7)	NS
(Global-Pt)	(D)	2.4 (0.6)	NS	1.9 (0.6)	
Analgesic consumption	(M)	1.8 (2.0)	NS	1.2 (1.9)	NS
(daily)	(D)	1.3 (1.8)	NS	1.0 (1.8)	

* Meclomen.
† Diclofenac.

between group differences were noted in body temperature, systolic blood pressure, diastolic blood pressure, hematology, clinical chemistry or urinalysis. With the exception of a significant increase ($p = 0.0005$) of BUN in the Meclomen group and a significant increase ($p = 0.0027$) in urine specific gravity with Voltaren, no within group alterations in hematology, clinical chemistry or urinalysis were noted on either drug.

RE. The RE of WOMAC versus Doyle and Lequesne Indices was calculated by combining data from the 2 treatment groups and is illustrated in Table 5. Although a total score can be developed for the Doyle and Lequesne indices, we have not previously attempted to aggregate the 3 subscales of WOMAC since their relative clinical importance has yet to be determined. However, to compare total index scores we have performed 2 types of aggregation in this analysis. The first was simple addition of the 3 subscale scores to give a total score. The total WOMAC score respects severity of symptoms but neglects both the relative clinical importance of the subscales and the difference in their scale lengths. The second aggregation was complex and involved combining subscale scores to give a pooled (P) score using the technique previously reported by Smythe, *et al*[20]. The pooled WOMAC score is based on derived standard deviation units and respects the severity of symptoms as well as the differ-

Table 4. *Adverse reactions reported by the 2 treatment groups*

Reactions	Diclofenac (n = 31)	Meclomen (n = 30)
Constipation	2	0
Depression (lassitude)	0	1
Diarrhea	2	3
Dyspepsia	0	2
Epigastric pain (low abdominal cramps)	0	1
Fluid retention (fingers/feet swollen)	1	0
Frequent bowel movement	0	1
Headache	2	1
Heartburn and gas	2	0
Hemorrhoidal bleeding	2	0
High blood pressure	0	1
Indigestion	0	1
Malaise	1	0
Migraine	0	1
Nausea	2	1
Rash	0	1
Shortness of breath	1	0
Stomach burning	0	4
Stomach cramps	0	2
Total number of adverse reactions (ADR)†	15	20
Total number of patients experiencing ADR	7	8
Total number of patients with severe ADR	1	2
Total number of patients withdrawn due to ADR	2	5
Total number of patients withdrawn due to inefficacy	1	1

† Some patients had more than one ADR.

Table 5. *Comparison of relative efficiency of WOMAC, Lequesne and Doyle indices*

WOMAC	Doyle	Lequesne			
		Pain	Maximum Distance Walked	Activities of Daily Living	Total
Pain	1.09	0.81	1.54	1.04	0.68
Stiffness	0.87	0.64	1.26	0.83	0.55
Physical function	1.17	0.86	1.65	1.14	0.73
Total (simple)*	1.32	0.94	1.81	1.22	0.82
Pooled (complex)	1.37	0.99	1.91	1.29	0.89

* The total WOMAC index score is calculated by simple addition of the 3 component subscale scores. The pooled WOMAC index score is calculated by a complex formula based on standard deviation unit and has been employed by Smythe, *et al*[19] in the pooled index.

ence in scale length of the 3 subscales, although it still does not account for any differences in their relative importance. RE values for both the total and pooled WOMAC versus the Doyle and Lequesne indices tended to be higher (range = 0.82–1.91) than the RE values for the 3 individual WOMAC subscales and those same indices (range = 0.55–1.65). Nevertheless, the range of RE values observed was small (0.55–1.91), none being exactly unity. Overall, the RE of the WOMAC, Doyle and Lequesne indices was similar.

Sample size calculation. From this study we calculated standard deviation values for estimating sample size for future trials using the VA 3.0 version of WOMAC. The calculated values are different for studies based on change scores than absolute or static scores. The recommended standard deviation values are as follows: For calculations based on absolute scores, Pain = 107.13, Stiffness = 48.16, Physical Function = 351.03. For calculations based on change scores, Pain = 91.06, Stiffness = 47.48, Physical Function = 199.72. The percent changes in mean values detected in our study were as follows: Meclomen, Pain = 42%, Stiffness = 43%, Physical Function = 40%; Voltaren, Pain = 19%, Stiffness = 19%, Physical Function = 19%.

DISCUSSION

The WOMAC Osteoarthritis Index has been validated with respect to reliability, face, content, construct validity and responsiveness, both in the context of an NSAID trial and a total joint arthroplasty study[4,5]. In those studies, the full version of LK3.0 (and LK3.0S) was used but the VA3.0 (and VA3.0S) version was only partially replicated owing to the large number of questions posed for purposes of construct validation. Our study, therefore, represents the first formal application of the full WOMAC VA 3.0 version. As well as permitting the comparison of Meclomen 100 mg po tid and Voltaren 25 mg po tid, our study also allowed additional characteristics of the WOMAC Index to be explored. In particular, we examined the relative efficiency of the instrument against 2 other validated indices and generated better estimates for the standard deviation of WOMAC subscales both for studies using static scores as well as those based on change scores.

In our study, in spite of randomization, the groups differed on WOMAC subscale scores although they did not differ in age, sex or disease duration. This can be considered the random result of the randomization process. The randomization process does not guarantee group similarity, but merely increases the probability that the 2 treatment groups will not differ. To adjust for the baseline imbalance, we employed an analysis of covariance technique. This is the traditional statistical approach used to address imbalance in efficacy measures. In our study it is considered appropriate since tests of homogeneity of variance between centers and treatments, based on demographic and efficacy measures, revealed no extreme departures, i.e., variance was homogeneous.

With respect to drug efficacy, all 3 WOMAC subscales detected clinically important statistically significant improvements on both drugs. The comparative analysis indicated a superiority of Meclomen 100 mg tid over Voltaren 25 mg tid for pain and stiffness, no difference being detected for physical function. The superiority of Meclomen 100 mg tid over Voltaren 25 mg tid was also verified by study joint measures of night pain and pain on walking and further confirmed by the pain/discomfort and activites of daily living subscales of the Lequesne Index. It should be noted that we have not compared Meclomen 100 mg tid with higher doses of Voltaren (i.e., 150 mg/day), and it may be that with the higher dosage, the 2 drugs are equal in potency. Overall the data indicate that Meclomen 100 mg tid is an efficacious NSAID in the treatment of OA of the knee and is superior to Voltaren 25 mg tid.

The percentage of patients experiencing adverse reactions was no different between the 2 drugs and was similar to that in many other NSAID trials in OA[8]. Although there were numerically more adverse reactions, more patients experiencing adverse reactions, more severe adverse reactions and more withdrawals due to adverse reactions with Meclomen than Voltaren, these differences were not statistically significant. Furthermore, the use of a relatively low dose of Voltaren may have restricted the frequency and severity of Voltaren related adverse reactions. At a higher dose (i.e., Voltaren 150 mg/day), a greater number of adverse reactions might be expected, some of which might be severe and

result in withdrawal. Qualitatively, the profile of adverse reactions was predictable, since gastrointestinal reactions are characteristic of the NSAID class of antirheumatic drugs. No significant decline in hemoglobin was noted with either NSAID. It is of note that while elevation of the BUN occurred with Meclomen, there was no corresponding elevation of the serum creatinine and no patient was withdrawn due to renal impairment. Furthermore, there was no clinically important or statistically significant difference between the 2 drugs with respect to vital signs (temperature, blood pressure), or other laboratory variables (hematology, biochemistry and urinalysis). We do not feel that the increase in urine specific gravity with Voltaren was clinically important. Overall we conclude that the tolerability profile of Meclomen is similar to that of Voltaren.

The relative efficiency statistic is one measure of the comparative responsiveness of different measures. However, statistical efficiency is not the only clinimetric property of importance. Indeed, as the content validity or comprehensiveness of a measurement process improves, so does its sample size requirements. This occurs by virtue of the inclusion of additional relevant but poorly responsive, items. The RE value is rarely unity and yet there are no standards for defining the significance of values < 1.0 or > 1.0. It should be noted that estimates of RE are subject to some variability and may differ between studies using the same instruments. Finally, it should be noted that the method of squaring the ratio of 2 t values results in a potential range of RE values that shows nonlinear progression. Finally, an RE value of, for example, 2.0 does not indicate that the numerator index is twice as efficient as the denominator index[19]. In comparing the WOMAC, Lequesne and Doyle indices, RE values were in general close to unity (0.55–1.91). We do not currently advocate the summation of the WOMAC subscales into a total or pooled index since complex issues of weighting and aggregation have not been resolved, although are the subject of current study. Nevertheless, it does facilitate the comparison of WOMAC with other indices in which aggregation is performed. We conclude from these data that the WOMAC, Doyle and Lequesne indices are similar in their RE. Although similar in statistical efficiency the 3 indices differ conceptually. The Doyle Index[14] only assesses articular tenderness and is subject to interobserver variability. Since it records joint disease in all target areas for OA it is not specific for knee disease and may detect improvements or deteriorations in other joint areas. For an OA knee study it collects unnecessary data and does not by itself have sufficient content validity (comprehensiveness) to act as the sole outcome measure. The Lequesne Knee Index[13] measures several important aspects of OA (pain, stiffness, walking distance and 4 forms of physical activity). However, neither its pain or stiffness nor its physical activity inventory is as comprehensive as that of WOMAC. Furthermore, its clinimetric properties have been less extensively documented. In contrast, WOMAC has a comprehensive inventory of questions developed by 100 patients with OA themselves[3]. The face, content and construct validity, test retest reliability, internal consistency and responsiveness have been determined for both Likert and VA scaled versions of the index[4,5]. We have also examined issues of parametric versus nonparametric analysis of the data[4,5], signal versus nonsignal approaches to measurement[7], blind versus informed administrations[21], the relationship between the severity of involvement and the corresponding clinical importance of discomfort or disability on each question[22] and the effect of circadian variation on perceived pain[23]. WOMAC is a high performance self-assessment outcome measure which assesses most of the important clinical consequences of OA knee and has been extensively validated in several different clinical settings (NSAID trials[5,24], total joint arthroplasty[4,25], interferential current therapy[26]).

In general it is difficult to find accurate values for the delta and standard deviation of a measure on which to calculate sample size for clinical trials. Type II errors may account for some of the lack of between drug differences detected in comparative studies in OA in the last 20 years[8]. Data from our study allow the calculation of WOMAC subscale standard deviations, not only for studies using absolute or static scores, but also those using change scores. These standard deviations are based on pooled estimates from the 61 participating patients. The selection of delta is an individual decision based on the research hypothesis being tested, the types of patients under study and the nature of the intervention[27]. However, the mean values at baseline and termination, as well as mean change scores, when considered in the context of the appropriate standard deviation, permit the calculation of sample size by future users of WOMAC. French and Swedish versions of WOMAC have been produced, the index having been applied in a variety of settings including studies of other NSAID, total hip arthroplasty and interferential current therapy. To date it has proven a valid, reliable and responsive self-administered measure in patients with OA of the knee and/or hip.

This comparison of Meclomen and Voltaren achieved 4 major objectives. We demonstrated in this group of patients that Meclomen 100 mg tid is a safe and efficacious NSAID, superior in efficacy and similar in tolerability to Voltaren 25 mg tid. It should be noted that we have not compared both drugs at maximum dosage and therefore cannot make any comprehensive statement regarding comparative drug efficacy or tolerability. Indeed, lack of comparability in doses of Meclomen and Voltaren may account for both benefit and ADR differences noted. The lack of group comparability in outcome measures at baseline may have introduced a bias that could possibly operate in favor of Meclomen. However no clinical correction can be made for this but a traditional statistical correction has been applied that should adequately, if not entirely, deal with the problem. The data, notably the

within group comparisons, suggest that Meclomen is a useful agent in the treatment of OA knee patients. Furthermore, we have determined that WOMAC shares a similar statistical efficiency with 2 other validated indices (Lequesne-knee and Doyle), and can specify the standard deviation necessary for calculating sample size for future OA studies using WOMAC as the principal outcome measure.

REFERENCES

1. Bellamy N, Buchanan WW: Outcome measurement in osteoarthritis clinical trials: the case for standardisation. *Clin Rheumatol 1984;*3:293–303.
2. Bellamy N: Osteoarthritis — An evaluation index for clinical trials. MSc Thesis, 1982, McMaster University, Hamilton, Canada.
3. Bellamy N, Buchanan WW: A preliminary evaluation of the dimensionality and clinical importance of pain and disability in osteoarthritis of the hip and knee. *Clin Rheumatol 1986;*5:231–41.
4. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt L: Validation study of WOMAC: a health status instrument for measuring clinically-important patient-relevant outcomes following total hip or knee arthroplasty in osteoarthritis. *J Ortho Rheumatol 1988;*1:95–108.
5. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW: Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol 1988;*15:1833–40.
6. Bellamy N: Pain assessment in osteoarthritis: experience with the WOMAC Osteoarthritis Index. *Semin Arthritis Rheum 1989;*(suppl 2)18:14–7.
7. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Duku E: Signal measurement strategies: are they feasible and do they offer any advantage in outcome measurement in osteoarthritis? *Arthritis Rheum 1990;*33:739–45.
8. Rosenbloom D, Brooks PM, Bellamy N, Buchanan WW: *Clinical Trials In The Rheumatic Diseases — A Selected Critical Review.* New York: Praeger Scientific, 1985:19–69.
9. Dresner AJ: Multicenter studies with sodium meclofenamate (Meclomen®) in the United States and Canada. *Curr Ther Res 1978;*(suppl)23:S90–106.
10. Willkens R: Experiences in the treatment of osteoarthritis of the hip and knee with sodium meclofenamate (Meclomen®). *Curr Ther Res 1978;*(suppl)23:S81–9.
11. Petrick TJ, Bovenkerk WE: Multicenter studies in the United States and Canada of meclofenamate sodium in osteoarthritis of the hip and knee. Double-blind comparison with placebo and long-term experience. *Arzneimittelforschung 1983;*(4A)33:644–8.
12. *Atlas of Standard Radiographs — The Epidemiology of Chronic Rheumatism.* Vol. 2. Philadelphia: F.A. Davis , 1963.
13. Lequesne MG, Mery C, Samson M, Gerard P: Indexes of severity for osteoarthritis of the hip and knee: Validation-Value in comparison with other assessment tests. *Scand J Rheumatol 1987;*(suppl)65:85–9.
14. Doyle DV, Dieppe PA, Scott J, Huskisson EC: An articular index for the assessment of osteoarthritis. *Ann Rheum Dis 1987;*40:75–8.
15. *SIR/DBMS (Scientific Information Retrieval/Database Management System), Version 2.2.18.* Deerfield: SIR (USA) Inc., 1988.
16. *SAS/STAT®, Version 6.03.* Cary: SAS Institute Inc., 1987.
17. *BMDP-UX Statistical Software.* BMDP Statistical Software Inc. Regents of the University of California, Los Angeles, CA. *Conversion HP-UX,* Innovus Inc.: Hamilton, 1988.
18. Feinstein AR: Scientific decisions for data and hypotheses. In: *Clinical Epidemiology — The Architecture of Clinical Research.* Philadelphia: WB Saunders, 1985;515–7.
19. Liang MH, Larson MG, Cullen KB, Schwartz JA: Comparative measurement efficacy and sensitivity of five health status instruments for arthritis research. *Arthritis Rheum 1985;*28:542–7.
20. Smythe HA, Helewa A, Goldsmith CH: "Independent Assessor" and "Pooled Index" as techniques for measuring treatment effects in rheumatoid arthritis. *J Rheumatol 1977;*4:144–52.
21. Bellamy N, Goldsmith CH, Buchanan WW, Campbell J, Duku E: Prior score availability: observations using the WOMAC Osteoarthritis Index. *Br J Rheumatol 1991;*30:150–60.
22. Bellamy N, Wells G, Campbell J: Relationship between severity and clinical importance of symptoms in osteoarthritis. *Clin Rheumatol 1991;*10(2):138–43.
23. Bellamy N, Sothern RB, Campbell J: Rhythmic variations in pain perception in osteoarthritis of the knee. *J Rheumatol 1990;*17:364–72.
24. Gerecz-Simon E, Bellamy N, Kean WF, Buchanan WW, Campbell J: Double-blind randomised controlled trial (rct) of sodium meclofenamate (meclomen) and diclofenac sodium (voltaren): post-validation reapplication of the WOMAC Osteoarthritis Index. XIIth European Congress of Rheumatology. In: Hodinka L, ed. *Hungarian Rheumatology, Volume XXXII 1991 Supplementum,* P4-316, p.364. Budapest, June 30–July 6, 1991.
25. Laupacis A, Leslie K, Wong C, *et al:* A randomized double-blind study of cemented versus non-cemented hip protheses. Society for Clinical Trials 10th Annual Meeting, Minneapolis, May 1989.
26. Young SL, Woodbury MG, Fryday-Field K, Donovan T, Bellamy N, Haddad R: Efficacy of interferential current stimulation alone for pain reduction in patients with osteoarthritis of the knee: a randomized placebo control clinical trial. American Physical Therapy Association Annual Conference Boston, June 23–27, 1991. *Phys Ther 1991;*(suppl)71:52.
27. Bellamy N, Buchanan WW: Clinical evaluation in rheumatic diseases. In: McCarty DJ, ed. *Arthritis and Allied Conditions — A Textbook of Rheumatology,* 11th ed. Philadelphia: Lea & Febiger. 1989:158–86.

A Multicenter Study of Tenoxicam and Diclofenac in Patients with Osteoarthritis of the Knee

NICHOLAS BELLAMY, W. WATSON BUCHANAN, ANDREW CHALMERS, PETER M. FORD, WALTER F. KEAN, GUNNAR R. KRAAG, ERIKA GERECZ-SIMON, and JANE CAMPBELL

ABSTRACT. Objective. **To conduct the first Canadian study of the comparative efficacy and safety of tenoxicam and diclofenac in patients with primary osteoarthritis (OA) of the knee.**

Methods. **Tenoxicam 20 mg per os once daily (po od) was compared to diclofenac (Voltaren™) 50 mg per os 3 times a day (po tid) in a 12-week, double blind, randomized, controlled, multicenter, parallel trial. The primary outcome measure was the pain dimension of the WOMAC OA Index. Following an initial screening visit and a 3 to 7 day NSAID-free washout period (i.e., baseline), patients were assessed at Weeks 2, 4 and 12; assessments including some 15 efficacy variables and safety variables.**

Results. **Ninety-eight patients [tenoxicam (n = 48), diclofenac (n = 50)] participated in the trial. Statistically significant ($p \leq 0.05$) improvements in all 3 dimensions of the WOMAC OA Index and six efficacy variables were noted in both treatment groups. No significant between drug differences were noted on any efficacy variable. Significantly fewer patients reported adverse events in the tenoxicam group (21 vs 33, $p = 0.03$).**

Conclusion. **Tenoxicam is efficacious and well tolerated in patients with OA of the knee. In this group of patients it was similar in efficacy and superior in tolerability to diclofenac 150 mg/day (50 mg tid). Thus the benefit/risk ratio of tenoxicam was superior to that of diclofenac in this study. (*J Rheumatol 1993;20:999–1004*)**

Key Indexing Terms:
NSAID WOMAC OA INDEX CLINICAL TRIAL

Nonsteroidal antiinflammatory drugs (NSAID) are one of the most frequently prescribed drugs in the treatment of osteoarthritis (OA). Tenoxicam is a new oxicam compound which has been shown to exert antiinflammatory, analgesic and antipyretic activities, and to inhibit platelet aggregation, prostaglandin and thromboxane synthesis[1]. At an oral dose of 20 mg/day, the peak plasma concentration is reached within 1 to 4 h with a half-life of about 70 h[2]. After multiple dose regimens, steady state conditions are reached 10 to 14 days after initiation of a 20 mg/day oral dose regimen[3]. The optimal daily dose of tenoxicam has been shown to be 20 mg/day[4].

Our objective was to conduct the first Canadian study of the comparative efficacy and safety of tenoxicam and diclofenac (Voltaren™) in patients with OA under double blind conditions for a period of 12 weeks. Superior clinical efficacy to placebo has been demonstrated with both these NSAID[5-9]. In particular, we have applied the validated WOMAC Osteoarthritis Index[10,11] to assess comparative drug efficacy.

MATERIALS AND METHODS

Ninety-eight consecutive consenting outpatients who were considered by the investigators to have primary OA of the knee entered the study at 6 centers across Canada. The following inclusion and exclusion criteria were applied: Inclusion — Symptomatic primary OA of at least one knee for at least 3 months requiring treatment with NSAID, age 18–75 years; radiographic evidence (within last 6 months) consisting of narrowing of joint space, sclerosis, marginal lipping, bone cysts, osteophyte formation, with a minimum Grade 2 and maximum Grade 3 severity[12]; and written informed consent; Exclusion — Any other type of arthritis; pregnancy or lactation; hypersensitivity to salicylates, oxicams or other NSAID, use of oral anticoagulants, systemic or intraarticular corticosteroid use in the past 2 months; ARA Class IV functional capacity; clinically significant cardiovascular, gastrointestinal (GI), hepatic, renal or hematological disease; aspartate aminotransferase (AST) or serum creatinine greater than 10% above upper limits of reference range; history of alcohol or drug abuse; collagen disease; concomitant diseases, such as psoriasis, syphilitic neuropathy, ochronosis, metabolic bone disease, acute trauma with or without degenerative joint disease, which may affect joints.

From the Department of Medicine, The University of Western Ontario, London, McMaster University, Hamilton, Queen's University, Kingston, ON, University of British Columbia, Vancouver, BC, and the University of Ottawa, Ottawa, ON, Canada.

Supported by a grant from Hoffmann-La Roche Limited.

N. Bellamy, MD, MSc, FRCP (Glas, Edin), FRCPC, FACP, Professor of Medicine, Epidemiology and Biostatistics, University of Western Ontario; W.W. Buchanan, MD, FRFPS, FRCP (Glas, Edin), FRCPC, FACP, Professor, Department of Medicine, McMaster University; A. Chalmers, MD, FRCPC, Associate Professor, Department of Medicine, University of British Columbia; P.M. Ford, MB, LRCP (Lond), MRCP (Lond), MRCS (Eng), FRCPC, Professor, Queen's University; W.F. Kean, MD, FRCPC, FACP, FARA, Associate Clinical Professor, McMaster University; G.R. Kraag, MD, FRCPC, Professor, University of Ottawa; E. Gerecz-Simon, MD, Co-Director, Musculoskeletal Pain Research, Service of Rheumatology, McMaster University; J. Campbell, BA, Research Assistant, Department of Medicine, University of Western Ontario.

Address reprint requests to Dr. N. Bellamy, Division of Rheumatology, Suite 303, Colborne 3, Victoria Hospital, 375 South Street, London, ON, Canada N6A 4G5.

Submitted December 10, 1991 revision accepted November 10, 1992.

Following initial screening (medical history and examination), there was a 3 to 7 day NSAID-free washout period during which patients were instructed to cease all NSAID therapy and take monitored quantities of acetaminophen, 325 mg if necessary. Patients were then reassessed (baseline), and only those patients who experienced a deterioration during the washout period and overall increase in disease activity thereafter, were randomized to receive either one capsule of tenoxicam 20 mg and 2 placebo capsules or 3 50 mg diclofenac capsules daily. Randomization occurred in blocks of 4 within each center, and treatment by center interactions were done in every analysis. Study medications were taken at meal times in the morning, noon and night. The color, size and markings of all of the capsules were identical. Both assessors and patients were blind to the allocation until after completion of the study. During the active treatment phase, patients were permitted to take supplementary analgesia sparingly, if required, with acetaminophen (supplied by the investigator) up to 6 × 325 mg/day. Compliance to study medication was verified by pill counting and analgesic ingestion was also assessed from pill counts.

Patients were assessed at 5 points during the trial: Screening, 0 (Baseline), 2, 4 and 12 (termination) weeks. The following efficacy variables were measured at Weeks 0, 2, 4 and 12: (1) WOMAC-VA3.0 OA Index[13] (a 24-item self-administered questionnaire probing pain, stiffness and physical function). It should be noted that, although the WOMAC OA Index was completed for both knees, the statistical analysis was based only on the worst knee; (2) Doyle Articular Index[14] (a modification of the Ritchie Index in which 48 joints or joint units are assessed by palpation for tenderness on a scale of 0 – no tenderness, 1+ patient complains of pain, 2+ patient complains of pain and winces, 3+ patient complains of pain, winces and withdraws); (3) Swelling of the knee joints was assessed using a 4-point scale, i.e., 0 = none, 1 = no swelling visible but suggested on palpation, 2 = swelling visible and perceived on palpation, and 3 = obvious swelling causing tightened skin over the joint; (4) Articular function measured by knee flexion/extension with a goniometer (key joint); (5) 50-foot walk time; (6) Duration of morning stiffness (min) in the past 2–3 days; (7) Duration of inactivity stiffness (min), i.e., after sitting, lying or resting later in the day; (8) Pain at night in key joint [100 mm visual analog scale (VAS)]; (9) Pain on initiation of walking in key joint (100 mm VAS); (10) Pain on walking in key joint (100 mm VAS); (11) Pain after a day of normal activity in key joint (100 mm VAS); (12) Severity of general pain (100 mm VAS); (13) Global assessment by patient (improved, no change, worsened); (14) Global assessment by physician (improved, no change, worsened); and (15) Physician and patient global evaluation of the clinical efficacy of the treatment at study termination compared to baseline (markedly improved, moderately improved, slightly improved, no change, slightly worse, moderately worse, markedly worse). At the screening visit, one knee was identified as the key joint[15] (i.e., the worst knee was the primary focus of measurement for future assessment in the study). All 6 of the investigators conducting the clinical assessment were standardized prior to the study[16]. A decision had been made *a priori* to use the WOMAC OA Index pain subscale as the principal outcome measure, all other variables being relegated to secondary outcome status.

Safety variables included laboratory determinations (performed in each investigator's regional laboratory) at screening, Week 4 and termination (i.e., routine hematology, clinical chemistry, urinalysis, and occult blood) and clinical determinations (i.e., blood pressure, heart rate and weight). Records were kept of concomitant medications. Adverse reactions were ascertained by open ended indirect questioning of the patient. Tolerance was evaluated both by the patient (very good, good, fair, poor, very poor) and the physician (based on the severity and frequency of side effects).

Sample size calculation for this study was based on the WOMAC pain subscale[17], the data being derived from a previous NSAID study[18] (Δ = 25% of 240.984, SD = 107.126). This permitted the detection of the clinically important between drug differences between active drugs with Type I and Type II error rates of $\alpha = 0.05_{2\text{-tailed}}$, $\beta = 0.20$ respectively with a sample size of 50 patients/group. A *post hoc* power analysis for the WOMAC pain subscale using the between group mean square error term and 3 repeated measures showed that a clinically important change of 20% from baseline could be detected with power of 0.99.

Following study completion, double data entry was performed by the sponsor, Hoffmann-La Roche Limited, using a SIR (Scientific Information Retrieval)[19] database. Data analysis, performed in collaboration with Innovus Inc., Hamilton, ON, was accomplished using SAS Version 6.03[20] and BMDP-UX[21] (1988) software packages. Data were checked for validity, normality, skewness and kurtosis. Descriptive analyses were conducted on all parametric variables defining means and standard deviations. Baseline comparability of demographic and efficacy variables was conducted. Continuous variables were analyzed by repeated measures analysis of covariance using the respective baseline scores (Week 0) as the covariates to compare the 2 treatment groups and six centers.

Compliance data and acetaminophen use were analyzed by repeated measures analysis of variance, comparing treatment groups and centers at Weeks 2, 4 and 12. All repeated measures analyses were adjusted for unequal time between measures. Categorical outcome measures were analyzed using Pearson χ^2 analysis to compare treatment groups and McNemar's χ^2 analysis was used to compare treatment groups and visits. Categories for the swelling score and global assessments which had low frequencies were collapsed for analysis. In all statistical tests, the level of Type I error (2-tailed) was set at 0.05. We made no correction for multiple comparisons, justified on the basis that a single primary efficacy variable was defined *a priori*.

RESULTS

One hundred patients entered the trial of whom 2 were excluded from analysis for the following reasons: allergy to aspirin = 1; allergy to acetaminophen = 1. Of the remaining 98 patients, 48 received tenoxicam and 50 received diclofenac. Included in all analyses, on an intent-to-treat basis, were 2 patients who committed protocol violations: one patient in the tenoxicam group had a history of alcohol or drug abuse and had a baseline ASOT or creatinine level greater than 10% above the upper limit of the reference range, and one patient in the diclofenac group was over 75 years of age. The disease and demographic profiles of patients are illustrated in Table 1. No clinically important between group differences were noted for any of these variables. The mean values of the efficacy outcome measures are illustrated in Table 2 at each of the clinical assessments. The number of patients completing therapy and the reasons for discontinuation are reported in Table 3.

Table 1. *Prerandomization, preintervention comparison of treatment groups' mean (standard deviation)*

Variable	Tenoxicam (n = 48)	Diclofenac (n = 50)
Sex (F, M)	34F, 14M	36F, 14M
Age (years)	62.5 (7.6)	62.7 (8.6)
Disease duration (years)	10.4 (8.9)	8.2 (7.6)
Definite osteophytes	34	40
Moderate multiple osteophytes	17	18
ARA Functional Class I	12	11
II	29	31
III	7	7
Weight (kg)	85.1 (17.6)	86.6 (18.6)
Height (cm)	166.2 (9.9)	166.9 (8.7)
Heart rate (bpm)	72.0 (9.5)	71.0 (10.3)
Systolic blood pressure (mm Hg)	139.6 (18.4)	140.7 (17.5)
Diastolic blood pressure (mm Hg)	83.5 (8.5)	82.5 (6.9)

Table 2. *Mean values for efficacy outcome measures at baseline (standard deviation) and visits at Weeks 2, 4 and 12*

Outcome Variable	Group	Baseline	Week 2	Week 4	Week 12	Within Group Baseline vs Termination p Value	Between Group Baseline vs Termination p Value
WOMAC OA Index:							
Pain (0–500 mm)	T	230.9 (95.5)	195.4	179.7	188.8	0.0372	0.7560
	D	231.7 (108.0)	177.3	146.3	157.4	0.0001	
Stiffness (0–200 mm)	T	97.4 (48.3)	75.6	65.6	71.0	0.0005	0.7447
	D	91.2 (47.4)	68.5	56.5	51.5	0.0001	
Physical function (0–1700 mm)	T	809.2 (361.2)	664.6	594.0	600.3	0.0005	0.4697
	D	773.9 (388.2)	585.6	494.9	503.2	0.0001	
Doyle articular index tenderness score (0–144)	T	13.8 (14.4)	11.0	8.9	9.0	0.0140	0.8899
	D	14.8 (13.6)	10.0	8.7	8.6	0.0029	
Pain severity (100 mm)	T	60.0 (20.5)	45.3	45.0	48.8	0.0339	0.5616
	D	57.8 (24.1)	43.8	39.7	43.3	0.0292	
Pain at night (100 mm)	T	45.2 (28.0)	35.5	34.0	36.4	0.0517	0.8670
	D	47.7 (27.7)	36.5	33.1	30.3	0.0006	
Pain on initiation of walking (100 mm)	T	60.0 (24.1)	44.0	43.3	35.9	0.0001	0.7196
	D	49.0 (24.7)	37.9	33.2	30.0	0.0001	
Pain on walking (100 mm)	T	56.4 (22.1)	47.2	42.1	38.4	0.0007	0.9987
	D	55.1 (24.2)	39.2	33.7	33.4	0.0001	
Pain after a normal day of activity (100 mm)	T	64.7 (22.5)	53.6	49.4	43.9	0.0001	0.8198
	D	59.9 (24.9)	46.0	37.3	38.9	0.0001	
Morning stiffness (min)	T	31.1 (45.7)	23.2	24.2	19.1	0.9652	0.2751
	D	21.0 (23.2)	19.8	15.5	14.6	0.3981	
Inactivity stiffness (min)	T	12.6 (15.0)	9.2	9.2	8.3	0.0650	0.5236
	D	14.3 (21.4)	7.8	7.8	13.3	0.0151	
50' walk time (s)	T	15.3 (7.1)	14.1	13.5	13.4	0.0047	0.7090
	D	13.5 (3.6)	12.9	12.1	12.2	0.0083	
Flexion (°)	T	115.2 (24.5)	117.6	119.2	122.5	0.0868	0.4332
	D	115.7 (19.5)	121.1	122.5	123.3	0.0003	
Extension (°)	T	1.4 (2.6)	1.3	1.2	1.9	0.3724	0.1954
	D	4.0 (6.1)	2.6	2.4	2.6	0.0318	
Total range (°)	T	114.1 (24.9)	120.3	118.0	121.0	0.1107	0.9044
	D	112.1 (20.7)	118.5	120.1	120.8	0.0001	
Body weight (kg)	T	85.6 (18.1)	86.5	85.6	84.0	0.7484	0.0252
	D	86.1 (18.5)	87.0	86.3	84.9	0.9312	
Heart rate (bpm)	T	70.7 (10.1)	71.1	72.7	72.2	0.1440	0.0780
	D	72.8 (11.0)	71.5	71.5	72.3	0.6663	
Systolic blood pressure (mm Hg)	T	134.5 (15.7)	133.9	138.5	135.7	0.6842	0.5643
	D	137.5 (15.2)	135.7	137.2	140.6	0.2226	
Diastolic blood pressure (mm Hg)	T	79.8 (9.6)	80.8	81.3	82.1	0.1063	0.7684
	D	80.6 (8.7)	81.2	81.5	82.3	0.4739	

T = Tenoxicam, D = Diclofenac.

Efficacy. Within group comparisons (baseline versus endpoint) were made for all outcome variables (Table 2). Statistically significant improvements with treatment were noted as follows: Both drugs – WOMAC (all 3 subscales), Doyle Articular Index, pain severity, pain on initiation of walking, pain on walking, pain after a normal day of activity, and 50-foot walk time; diclofenac only – pain at night, inactivity stiffness, flexion, extension and total range of movement. With respect to key joint swelling, 25% of the tenoxicam group and 20% of the diclofenac group had either obvious or visible swelling at baseline. Both treatment groups experienced a reduction from baseline in swelling over the course of the study. By the end of the study, the global assessments made by both the physician and patient showed a similar improvement in both groups, i.e., tenoxicam (physician 40%, patient 38%), diclofenac (physician and patient 34%). Physicians and patients rated global evaluation of clinical efficacy similarly at study termination (i.e., tenoxicam/diclofenac: improved markedly – 11,21%/33,39%; improved moderately – 38,30%/25,19%; improved slightly – 13,6%/18,14%; no change – 19,19%/14,10%; worsened slightly – 2,2%/4,4%; worsened moderately – 9,11%/6,8%, and worsened markedly 8,11%/0,6%. No significant treatment differences were found for any of these assessments.

Between treatment comparisons were made for all outcome variables. No statistically significant treatment differences were found between tenoxicam and diclofenac on any of the efficacy measures. There were no differences in analgesic

Table 3. *Number of patients completing therapy and reasons for discontinuation*

	Tenoxicam	Diclofenac	Total
Randomized	48	50	98
Completed 12 weeks' treatment	36	33	69
Discontinued therapy before end of Week 12[a]	11	16	27
Reason for discontinuation:			
Lack of efficacy[b]	8	2	10
Adverse event[c]	3	11	14
Concurrent illness	0	1	1
Medication missing (noncompliant)	0	1	1
Patient withdrew	0	1	1

[a] $\chi^2 = 1.02$, p value = 0.3140.
[b] $\chi^2 = 4.29$, p value = 0.0384.
[c] $\chi^2 = 4.96$, p value = 0.0259.

consumption. However, 2 patients in the diclofenac group and 8 in the tenoxicam group withdrew due to inefficacy ($\chi^2 = 4.29$, df = 1, p = 0.04) (Table 3).

Tolerability. The frequency and nature of adverse events reported during the study are illustrated in Table 4. The majority of clinical adverse events reported were gastrointestinal in nature and accounted for 40–44% of all events in both treatment groups. As shown in Table 4, 44% of patients in the tenoxicam group and 66% of patients in the diclofenac group reported one or more adverse event, and this rate significantly differed between the treatment groups ($\chi^2 = 4.90$, df = 1, p = 0.03). Likewise, there were numerically more adverse events reported in the diclofenac group (n = 85) than in the tenoxicam group (n = 52). No significant difference was detected in the number of patients with severe events (tenoxicam = 8, diclofenac = 13). There was, however, a significant difference in the number of patients withdrawn due to adverse events (tenoxicam = 3, diclofenac = 11, $\chi^2 = 4.96$, df = 1, p = 0.03). The reasons for withdrawal were as follows: tenoxicam — dyspepsia = 1, fluid retention = 1, rash = 1; diclofenac — fluid retention = 1, headache = 1, weight gain = 1, asthma = 1, gastrointestinal events = 5, elevated glucose, leg cramps and epigastric pain = 1, bowel strangulation, headache, abdominal pain = 1. At study termination, evaluation of tolerance by physicians and patients showed no significant treatment differences. The physicians' evaluation of adverse events was as follows, tenoxicam/diclofenac: no adverse events (AE) = 62/39%; mild AE = 4/15%; moderate AE = 21/27%; severe AE = 13/19%. The patients' evaluation of tolerance showed similarity between drugs, i.e., tenoxicam/diclofenac: very good = 60/45%; good = 23/21%; fair = 11/18%; poor = 0/4%; very poor = 6/12%. No significant between group differences were noted in heart rate, systolic blood pressure, diastolic blood pressure, hematology, clinical chemistry or urinalysis. The number of normal baseline and abnormal termination values are reported for all laboratory measures in Table 5. With the exception of a sig-

Table 4. *Adverse events by the 2 treatment groups*

Event	Tenoxicam (n = 48)	Diclofenac (n = 50)
GI system disorders	23	34
Central and peripheral nervous disorders	8	18
Body as a whole — general disorders[a]	7	8
Respiratory system disorders	5	5
Metabolic and nutritional disorders	5	5
Musculoskeletal disorders	0	6[b]
Psychiatric disorders	0	5[c]
Skin and appendages disorders	2	2
Urinary system disorders	1	1
Vascular disorders	1	0
Autonomic nervous system disorders	0	1
Total number of adverse events (AE)[†]	52	85
Total number of patients experiencing AE[‡]	21	33
Total number of patients with severe AE[‡]	8	13

[a] WHO Classification of Adverse Events (1990); i.e., fatigue, fever, cold, etc.
[b] Musculoskeletal: leg pain (2), back pain (1), shoulder and neck pain (1), foot pain (2).
[c] Psychiatric: depression (1), increased appetite (1), insomnia (3).
[†] Some patients had more than one AE.
[*] $\chi^2 = 4.901$, p value = 0.0270.
[‡] Side effects were graded on a 3-point scale (mild, moderate, severe) at the investigator's discretion; $\chi^2 = 0.000$, p value = 0.9886.

Table 5. *Safety laboratory measures: Number of normal baseline and abnormal termination values*

Laboratory Measure	Tenoxicam	Diclofenac
Hematology		
Hemoglobin	1	1
Hematocrit	2	0
Red blood cells	0	1
White blood cells	2	2
Platelets	1	0
Biochemistry		
Total protein	1	0
Albumin	1	0
Phosphorus	0	0
Glucose	1	4
Blood urea nitrogen	13	9
Creatinine	2	1
Total bilirubin	2	1
Alkaline phosphatase	1	2
Chloride	4	7
Potassium	0	1
Sodium	1	2
Aspartate aminotransferase	0	7
Calcium	0	4
Urinalysis		
Specific gravity	3	0
pH	2	1
Albumin	1	1
Glucose	0	0
Fecal occult blood	1	1

nificant increase in abnormal BUN at termination compared to baseline in both the tenoxicam (χ^2 = 10.29, df = 1, p = 0.001) and the diclofenac groups (χ^2 = 4.46, df = 1, p = 0.03), and significantly more abnormal SGOT at termination than at baseline in the diclofenac group (χ^2 = 7.00, df = 1, p = 0.01), no other within group changes in hematology, clinical chemistry or urinalysis were noted with either drug. Body weight was found to decrease significantly more in patients in the tenoxicam group than in patients in the diclofenac group (p = 0.03). Although this difference was statistically significant, the mean difference between groups was less than 2 kg.

DISCUSSION

The prescription of NSAID to patients with OA of the knee is a trade off between risk and benefit. Fortunately, for the majority of NSAID, the benefits are substantial and accrue to the majority of recipients, while adverse events are relatively infrequent. The prediction of response in individual patients is largely speculative[22] since we have an incomplete knowledge of those factors that determine the response[23] which is often mild in degree and usually either self-limited or easily managed. Nevertheless, comparative drug efficacy and tolerability, among other factors such as cost and convenience, are important indications in selecting individual drugs from those available for treating individual patients. Our study used a rigorous methodologic design and included a fully validated primary outcome measure (WOMAC) of defined reliability, face, content, construct validity and responsiveness[10,11] which has previously detected a clinically important and statistically significant difference between 2 NSAID[18]. WOMAC is currently available in Likert scaled and visual analog scaled forms, in English, French and Swedish language translations and has been used in several different clinical settings (pharmacologic[18,24,25], surgical[26], physiotherapy[27,28]). We are aware that it is presently being used in 12 other studies (5 pharmacologic, 4 surgical, 3 physiotherapy) in Canada, USA, Holland, Sweden and Australia. Many previous NSAID studies in OA have used outcome measures of poorly defined or undefined reliability, validity and responsiveness[29]. This has the disadvantages of lack of standardization of outcome measures and an inability to differentiate a Type II error from pharmacologic similarity of the test drugs.

In our study we observed beneficial pharmacologic effects with both tenoxicam and diclofenac but no significant difference between these 2 agents in their efficacy[30-32]. Given the use of a fully validated primary outcome measure and the well demonstrated efficacy of diclofenac[33], this suggests that tenoxicam 20 mg od is efficacious in the treatment of OA knee and similar in efficacy to diclofenac 150 mg/day (50 mg tid). This contention is supported by the majority of other efficacy outcome measures that also show significant within group improvements but no between group differences. It is particularly important to note that there was no between group difference in acetaminophen consumption indicating a lack of difference in analgesic effect between the 2 treatments. It is of interest that, while the WOMAC stiffness subscale (which assesses severity of morning stiffness and gelling) detected improvement with treatment, stiffness measures, based on duration of stiffness, failed to detect similar improvement except for inactivity stiffness in the diclofenac group. It is our impression that in patients with OA NSAID treatment more frequently reduces the severity of morning stiffness than its duration. This likely explains these apparently discrepant results. An alternate explanation is that the sample size required for measuring duration of stiffness effects may be higher than that used in our study. Our recent experience determining variables for sample size calculation in trial eligible patients with OA suggests that the minimum requirements for detecting a clinically important difference in the duration of morning stiffness are possibly 195–1368/group ($\alpha_{\text{1-tailed}}$ = 0.05, β = 0.20)[34].

The issue of tolerability is paramount in NSAID prescribing since in most published double blind studies, investigators have generally concluded that there is no between group difference between any 2 NSAID. In general, both tenoxicam and diclofenac were well tolerated by the majority of patients. However, tenoxicam may be significantly better tolerated as evidenced by the following: fewer patients reporting adverse events and withdrawals due to adverse events (p = 0.03), fewer GI events and fewer SGOT elevations. The superior GI tolerability of tenoxicam has been noted in other comparative trials[35,36], while a possible proclivity for diclofenac to produce hepatotoxicity has been the subject of a recent publication[37].

Clearly the efficacy and tolerability profile of an agent cannot be defined by a single trial. The conclusions, by necessity, are limited to patients of similar characteristics to those selected for study. Nevertheless, our trial provides additional data that support the contention that tenoxicam 20 mg po od is an efficacious agent similar in its effect to diclofenac 150 mg/day (50 mg tid). In this group of patients, tenoxicam was better tolerated particularly with respect to GI and hepatic (SGOT) events. Our data suggest that tenoxicam is an efficacious, well tolerated agent in the treatment of patients with OA of the knee. Its benefit/risk ratio was superior to that of diclofenac in our study.

REFERENCES

1. Bradshaw D, Cashin CH, Kennedy AJ, Roberts NA: Pharmacological and biochemical activities of "tenoxicam" (Ro 12–0068), a new nonsteroidal anti-inflammatory drug. *Agents Actions 1984;*15:569–77.
2. Heintz RC, Guentert TW, Evico JF, *et al:* Pharmacokinetics of tenoxicam in healthy human volunteers. *Eur J Rheumatol Inflamm 1984;*7:33–44.
3. Crevoisier C, Heizmann P, Forgo I, Dubach UC: Plasma tenoxicam concentrations after single and multiple oral doses. *Eur J Drug Metab Pharmacokinet 1989;*14:23–7.

4. Vischer TL: Efficacy and tolerability of tenoxicam — an overview. *Scand J Rheumatol 1987;*(suppl 65):107–12.
5. Merzario L, Piatti M: Tenoxicam, a new long-acting nonsteroidal anti-inflammatory drug. Comparative multicentre study versus placebo. In: Joint GS, ed. *Tenoxicam: Presentation to the EULAR Symposium 1986. Litera Rheumatological 1986.* Basel: EULAR Publishers, 1987;7:41–8.
6. Diverse P, De Moor M, Ginsberg FR, *et al:* Double-blind multicentre study of tenoxicam and piroxicam versus placebo in arthrosis of the knee and hip. In: Joint GS, ed. *Tenoxicam: Presentation to the EULAR Symposium 1986. Litera Rheumatologica 1986.* Basel: EULAR Publishers, 1987;7:19–27.
7. Amundsen T, Bieken L, Borkje B, *et al:* Variation in response to naproxen and diclofenac in patients with osteoarthritis. *Curr Therapeut Res 1983;*33:793–801.
8. Berry H, Bloom B, Hamilton EBD: A comparative study of piroxicam (Feldene), diclofenac and placebo in osteoarthritis. *Clin Trials J 1982;*19:349–58.
9. Germain PB: A placebo-controlled study of diclofenac sodium for the treatment of osteoarthritis of the hip and knee. *Curr Therapeut Res 1985;*37:259–68.
10. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt L: Validation study of WOMAC: a health status instrument for measuring clinically-important patient relevant outcomes following total hip or knee arthroplasty in osteoarthritis. *J Orthop Rheumatol 1988;*1:95–108.
11. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW: Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol 1988;*15:1833–40.
12. *Atlas of Standard Radiographs - The Epidemiology of Chronic Rheumatism.* Vol 2. Philadelphia: FA Davis, 1963.
13. Bellamy N: Osteoarthritis – An evaluative index for clinical trials. MSc Thesis, 1982, McMaster University, Hamilton, Canada.
14. Doyle DV, Dieppe PA, Scott J, Huskisson EC: An articular index for the assessment of osteoarthritis. *Ann Rheum Dis 1981;*40:75–8.
15. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Duku E: Signal measurement strategies: are they feasible and do they offer any advantage in outcome measurement in osteoarthritis. *Arthritis Rheum 1990;*33:739–45.
16. Bellamy N, Carette S, Ford PM, *et al:* Osteoarthritis antirheumatic drug trials. I. Effects of standardization procedures on observer dependent outcome measures. *J Rheumatol 1992;*19:436–43.
17. Colton T: Inference on means. In: *Statistics in Medicine.* Boston: Little, Brown, 1974:99–150.
18. Bellamy N, Kean WF, Buchanan WW, Gerecz-Simon E, Campbell J: Double blind randomized controlled trial of sodium meclofenamate (Meclomen) and diclofenac sodium (Voltaren): Post validation reapplication of the WOMAC Osteoarthritis Index. *J Rheumatol 1992;*19:153–9.
19. SIR/DBMS (Scientific Information Retrieval/Database Management System), Version 2.2.18, SIR(USA) Inc., 1988 Deerfield, Illinois.
20. SAS Institute Inc. SAS/STAT™, Version 6.03. Cary, NC: SAS Institute Inc., 1987.
21. BMDP-UX Statistical Software, BMDP Statistical Software Inc., Copyright® Regents of the University of California, Los Angeles, CA. Conversion HP-UX, Innovus Inc., Hamilton, Ontario, 1988.
22. Bellamy N: Prognosis in rheumatoid arthritis (editorial). *J Rheumatol 1991;*18:1277–9.
23. Bellamy N: *Prognosis in the Rheumatic Diseases.* Dordrecht/Boston/London: Kluwer Academic, 1991.
24. Bellamy N, Buchanan WW, Grace E: Double-blind randomized controlled trial of isoxicam vs piroxicam in elderly patients with osteoarthritis of the hip and knee. *Br J Clin Pharmacol 1986;*22:149S–55S.
25. Gerecz-Simon E, Bellamy N, Kean WF, Buchanan WW, Campbell J: Double-blind randomized controlled trial (RCT) of sodium meclofenamate (meclomen) and diclofenac sodium (voltaren): post-validation reapplication of the WOMAC Osteoarthritis (OA) Index. Hungarian Rheumatology. Abstracts of the XIIth European Congress of Rheumatology. Budapest, Hungary, June 30–July 6, 1991 Supplementum, Volume XXXII, Abstract P4–316, p.364.
26. Laupacis A, Leslie K, Wong C, *et al:* A randomized double-blind study of cemented versus non-cemented hip protheses. Society for Clinical Trials 10th Annual Meeting, Minneapolis, May 1989.
27. Young SL, Woodbury MG, Fryday-Field K, Donovan T, Bellamy N, Haddad R: Efficacy of interferential current stimulation alone for pain reduction in patients with osteoarthritis of the knee: a randomized placebo controlled clinical trial. American Physical Therapy Association Annual Conference, Boston, June 23–27, 1991. *Phys Ther 1991;*(suppl 6)71:R088.
28. Young SL, Woodbury MG, Fryday-Field K, Donovan T, Bellamy N, Haddad R: Efficacy of interferential current stimulation alone for pain reduction in patients with osteoarthritis of the knee: a randomized placebo controlled clinical trial (abstr). Canadian Physiotherapy Association, Montreal Meeting, May 20–24, 1991. *Physiotherapy Canada 1991;*43:4.
29. Bellamy N, Buchanan WW: Outcome measurement in osteoarthritis clinical trials: the case for standardization. *Clin Rheumatol 1984;*3:293–303.
30. Wille T, Scholz H-J: Comparative double-blind study of tenoxicam versus diclofenac retard in patients with osteoarthritis. In: Joint GS, ed: *Tenoxicam: Presentation to the EULAR Symposium 1986. Litera Rheumatologica 1986.* Basel: EULAR Publishers, 1987;7:62–8.
31. Tedesco P, Lion P: Comparative clinical study of a new non-steroidal anti-inflammatory drug — tenoxicam. 1987 XIth European Congress of Rheumatology. Athens. Poster Book No. 6. The Future Management of Rheumatic Diseases. Basle, March 2–3, 1990.
32. Dunky A, Fasching U, Eberl R, *et al:* Tenoxicam versus diclofenac retard in osteoarthritis — a double-blind, multicentre study. The 6th SEAPAL Congress of Rheumatology. Tokyo. Poster Book No. 18. The Future Management of Rheumatic Diseases. Basle, March 2–3, 1990. Roche.
33. Todd PA, Sorkin EM: Diclofenac sodium: A reappraisal of its pharmacodynamic and pharmacokinetic properties, and therapeutic efficacy. *Drug 1988;*35:244–85.
34. Bellamy N, Carette S, Ford PM, *et al:* Osteoarthritis antirheumatic drug trials. III. Setting the delta for clinical trials — results of a consensus development (Delphi) exercise. *J Rheumatol 1992;*19:451–7.
35. Dunky A, Fasching U, Eberl R, *et al:* Tenoxicam versus diclofenac retard in osteoarthritis of the knee. A double-blind multicentre study. *Drug Invest 1990;*(suppl 3)2:47–8.
36. Castillo A, Padrino JM, Roces A: Tenoxicam vs piroxicam in the treatment of gonarthrosis. A Poster presented at the EULAR Congress in Athens 1987.
37. Helfgott SM, Sandberg-Cook J, Zakim D, Nestler J: Diclofenac — associated hepatotoxicity. *JAMA 1990;*264:2660–2.

Double Blind Randomized Placebo Control Trial of Controlled Release Codeine in the Treatment of Osteoarthritis of the Hip or Knee

PAUL M. PELOSO, NICHOLAS BELLAMY, WILLIAM BENSEN, GLEN T.D. THOMSON, ZOLTAN HARSANYI, NAJIB BABUL, and ANDREW C. DARKE

ABSTRACT. Objective. Pain is the cardinal feature of osteoarthritis (OA), and with advancing disease there is loss of function and increasing pain even at times of joint rest. Few studies have evaluated the role of opioid analgesics in treating the pain of OA.

Methods. This randomized, double blind, parallel group study compared the efficacy and safety of a 12 hourly controlled release codeine formulation (Codeine Contin®) with placebo in patients with chronic pain due to OA of the hips and/or knees. The 4 week treatment period, following an analgesic washout phase of 2–7 days, included weekly clinic evaluations, at which the dose was escalated as appropriate, and daily patient diary completion. Pain (daily), stiffness, and physical function (weekly) were assessed using the multidimensional, self-administered WOMAC (visual analog scale version) questionnaire.

Results. Sixty-six eligible patients completed the study. The mean initial and final daily doses of controlled release codeine were 50 mg every 12 h at baseline and 159 mg every 12 h at the final assessment. All variables in the efficacy analysis indicated superiority of controlled release codeine over placebo. The WOMAC pain scale showed an improvement of 44.8% over baseline in the controlled release codeine group compared with 12.3% taking placebo ($p = 0.0004$). For the WOMAC stiffness and physical function scales the improvements over baseline on controlled release codeine were 47.7% and 49.3%, respectively compared with 17.0% and 17.0%, respectively, with placebo ($p = 0.003$; $p = 0.0007$). Controlled release codeine was also significantly better than placebo on measures of sleep quality and requirement for supplemental acetaminophen.

Conclusion. Single entity controlled release codeine is an effective treatment for pain due to OA of the hip or knee. (J Rheumatol 2000;27:764-71)

Key Indexing Terms:
CODEINE CONTROLLED RELEASE OSTEOARTHRITIS
RANDOMIZED CLINICAL TRIAL WOMAC

Osteoarthritis (OA) is a common clinical condition. Estimates suggest that about 52% of the population is affected, and in those over 65 years, up to 85% of individuals have some involvement. The prevalence of knee and hip OA increases progressively with age. Pain is the cardinal feature of clinical OA and is associated with the presence of radiographic changes and with increased mortality, morbidity, and functional dependence on others[1]. With advancing disease there is increasing pain and loss of function, with intrusion of pain even at times of joint rest.

Current American and Canadian recommendations[2-4] for the treatment of OA pain favor the use of simple analgesics, physical modalities, and topical therapies before nonsteroidal antiinflammatory drugs (NSAID). The relative efficacy of NSAID versus simple analgesics is being debated[5]. In addition NSAID can produce toxicity in the gastrointestinal tract[6-8] and both NSAID and acetaminophen can produce renal toxicity[9-12] that may offset their beneficial effect. The American and Canadian guidelines do not discuss the use of opioid analgesics in any detail.

In contrast, the role of opioids in chronic treatment of cancer associated pain has been well established, with adequate pain control being achievable in up to 85–95% of patients, without unmanageable or intolerable side effects and without significant risk of addiction[13,14]. However, application of these principles to the treatment of patients with unrelieved pain of non-cancer origin has been hampered by concerns over regulatory sanctions and the possibility that opioid side effects or

From the Royal University Hospital, Saskatoon, Saskatchewan; Victoria Hospital, London, Ontario; St. Joseph's Hospital/McMaster University, Hamilton, Ontario; St. Boniface Hospital, Winnipeg, Manitoba; and Department of Scientific Affairs, Purdue Frederick, Pickering, Ontario, Canada.

P.M. Peloso, MD, MSc, FRCPC, Assistant Professor of Medicine, University of Saskatchewan; N. Bellamy, MD, MSc, FRCPC, Victoria Hospital; W. Bensen, MD, FRCPC, St. Joseph's Hospital/McMaster University; G.T.D. Thomson, MD, FRCPC, St. Boniface Hospital; Z. Harsanyi, MBA, Department of Scientific Affairs, Purdue Frederick; N. Babul, PharmD, Royal University Hospital; A.C. Darke, PhD, Department of Scientific Affairs, Purdue Frederick.

Address reprint requests to Dr. P.M. Peloso, University of Saskatchewan, Room 369 Ellis Hall, Royal University Hospital, Saskatoon, Saskatchewan S7N OW8; E-mail:pelosop@duke.usask.ca

Submitted July 6, 1998 revision accepted July 7, 1999.

psychologic dependence will reduce patients' functional ability rather than improve it through reduction in pain[15]. There is a growing literature on the effectiveness of opioid analgesics in nonmalignant pain[16-18] and recently, controlled clinical trials have been completed indicating the effectiveness of controlled release morphine[19] and controlled release codeine[20,21]. This has led to the development of guidelines for the use of opioids in patients with pain of non-cancer origin[22,23].

However, few studies have looked specifically at the role of opioids in treating pain associated with OA despite its presumed nociceptive origin in multiple anatomic sites including capsule, ligaments, and insertions, and noncorpuscular type 4 nerve endings in the fibrous capsule, synovium, articular fat pads, and arterioles[24,25]. Recent demonstrations of the efficacy of intraarticular morphine in relieving pain after knee surgery, through an action on mu-opioid receptors[26-32], further support the rationale for trials of opioids in patients with joint pain.

While codeine is commonly used for the treatment of OA pain, usually in combination with acetaminophen or acetylsalicylic acid (ASA), there are no controlled studies demonstrating its benefits. Such studies are needed to balance clinical benefits against concerns over use in patients with noncancer pain. Regular administration of controlled release opioids results in improved compliance and pain control in patients with cancer pain[33]. The availability of a controlled release formulation of codeine with pharmacokinetics supportive of 12 hourly dosing[34] suggested that it may be of similar value in treatment of patients with OA pain. A previous study in patients with cancer pain revealed a dose–response relationship in the range of 200–600 mg/day[35]. To decrease the risk of opioid toxicity in patients with a lower level of prior exposure to opioids, we conducted a randomized placebo controlled trial of patients with OA. Patients began controlled release codeine at a daily dose of 100 mg with provision for titration to a maximum of 400 mg/day.

MATERIALS AND METHODS

Patient selection. Four Canadian sites participated after obtaining approval from an independent research ethics board. Patients with primary OA who were male or nonpregnant female and over the age of 35 years were eligible. Patients were required to have OA symptoms, including pain, stiffness, and disability, requiring the use of acetaminophen, antiinflammatory agents, or opioid analgesics for the previous 3 months or longer. Patients were required to provide written informed consent, and were given a copy of their signed consent. In all instances, patients had radiographic confirmation of minimum Grade II OA severity of a knee or hip joint, as defined by the standard atlas of radiographs[36]. Grade II OA changes require the presence of an osteophyte and joint space narrowing. Individuals with more advanced radiographic grades were also eligible provided that imminent joint replacement surgery would not prevent study completion. Patients were required to discontinue their prestudy analgesics, and experience a flare in their hip and/or knee pain during a 2–7 day washout period. A flare was defined as an increase in pain to a minimum report of moderate pain on a 5 point Likert scale (none, mild, moderate, severe, and excruciating pain).

Patients were excluded if they had a known allergy to codeine, other opioids, or acetaminophen, and if a history of previous opioid abuse, characterized by compulsive drug use, an intense desire for the drug, and an overwhelming craving for its continued availability, or if manipulation of a previous physician or the medical system for the purposes of obtaining additional drug was suspected. In addition, patients were ineligible if they had secondary causes of OA, or if they had received systemic or intraarticular corticosteroids in the past 2 months or intraarticular viscosupplementation in the past 6 months. Finally, patients with Grade 4 severity and awaiting replacement surgery were also ineligible.

Trial design. Medication at the 4 centers was dispensed according to a randomized, balanced, double blind parallel group assignment. Patients were allocated to either controlled release codeine (Codeine Contin®, Purdue Frederick, Pickering, Ontario, Canada) administered 12 hourly, or identical appearing placebo, also given 12 hourly. Use of additional antiinflammatory or analgesic medication, other than acetaminophen 650 mg up to 3 times daily for control of pain not managed by controlled release codeine or placebo, was not permitted. Acetaminophen use was recorded in a patient diary. Treatment was initiated at a dose of 100 mg of controlled release codeine per day (or identical placebo), and the dose was escalated weekly, provided there was ongoing pain and a lack of limiting side effects, up to a maximum of 400 mg per day (200 mg q12h). The treatment period after the washout phase was 4 weeks, with weekly clinical evaluations and daily completion of a diary by the patients.

Two primary measures of efficacy were established a priori. These measures were the daily Western Ontario and McMaster University Osteoarthritis (WOMAC) pain visual analog scale (VAS) and the daily overall Pain Intensity scores over the previous week. The WOMAC is a validated, multidimensional, self-administered questionnaire capable of measuring clinically important symptoms in patients with OA of the hip and/or knee, that has been validated in both VAS and Likert versions[37,38]. The WOMAC pain VAS consists of 5 questions on OA pain, including walking on a flat surface, going up or down stairs, at night in bed, at rest, and standing. The VAS version used in this study is a 100 mm scale, anchored by "no pain" at the zero mm anchor and "extreme pain" at the 100 mm anchor. The overall pain intensity over the previous week was assessed by asking the patient: "What was your average pain over the last week?"

Secondary measures included: weekly WOMAC Stiffness and Physical Function scales; daily 100 mm VAS average pain scale; 7 item questionnaire on sleep (4 items using 100 mm VAS scale: 1. trouble falling asleep; 2. need medications to fall asleep; 3. awakening by pain at night; and 4. awakening by pain in the morning); weekly nondirected adverse events questionnaire; and the Drug Liking Index completed at study termination. The Drug Liking Index is a 9 point Likert scale where 1 = "I dislike the drug effect very much" and 9 = "I like the drug effect very much"[39]. Patients were instructed to use this scale to assess the central nervous system effects of the drug, independent of its analgesic effects on their OA. Patient and physician global assessment of clinical effectiveness (a change score using a 7 point scale, where 0 = "markedly worse" and 7 = "markedly improved") was recorded at study termination.

Sample size and statistical evaluation. Sample size was determined for both primary outcomes. For the WOMAC pain VAS scale, the maximum score obtainable is 500 mm from five 100 mm scales. A previous study evaluating 2 NSAID showed a standard deviation (SD) of 107 mm, and a mean pain scale score of 241[40]. In order to detect a 100 mm difference between 2 therapies, a sample size of 25 patients per treatment arm is needed, when the power is set at 80% and the alpha is set at 0.05 (2 tailed). Allowing for a 25% dropout rate mandated a total study sample size of 68.

Although the WOMAC index has been used in a large number of clinical trials[1], it has not previously been used in an efficacy trial of an opioid analgesic. In addition to the multi-item pain measurement afforded by the WOMAC index, a second primary outcome measure based on a single global VAS pain scale was also used. Assessments of pain on a 100 mm VAS scale in studies in OA of the knee have shown an overall median SD of 23 mm[41] and it has previously been shown using a Delphi approach that a difference of 15 mm between treatments is clinically important[42]. Assuming a 20 mm difference between controlled release codeine and placebo to be a clinically important difference, and with alpha set at 0.05 (2 tailed) and power set at

80%, a total of 39 patients per group would be required. Again assuming a 25% dropout rate, a minimum of 104 patients across the study sites was calculated to be required.

Data were entered and analyzed in SAS v.6.12. For continuous and ordinal data, analysis was undertaken using means, SD, and analysis of covariance using baseline data as the covariate; baseline comparisons were made using the t test. Since rescue analgesic use was not assessed at baseline, analysis of variance was used for comparing treatments. Categorical variables were also compared using Fisher's exact test and the Wilcoxon test (Drug Liking Index). Differences across treatment visits were examined using repeated measures analysis of variance of the change from baseline scores and the differences between treatments at each visit were assessed by analysis of covariance. For descriptive purposes, values have been expressed as change from baseline and percentage change from baseline.

RESULTS

A total of 103 patients were enrolled, with 66 providing complete information across all measurement points. Fifty-one patients were initially randomized to controlled release codeine, and 52 to placebo. Of the patients completing the study 31 had been randomized to controlled release codeine and 35 to placebo. Seven patients in each group had previously used codeine on a longterm basis (a mean of 2.3 and 1.9 yrs in the controlled release codeine and placebo groups, respectively). As shown in Table 1, the proportion of patients who completed the study did not differ between treatments.

The mean initial and final controlled release codeine doses were 50 mg (SD 0.0) at baseline and 159 mg (SD 52) 12 hourly at the 4 week final assessment. The mean age of the completed patients was 61.6 (SD 11.2) years with a minimum of 39 years and a maximum of 81 years. There were 25 men (13 codeine, 12 placebo) and 41 women (18 codeine, 23 placebo). Hip pain was reported by 32 patients, 58 patients reported knee pain, and 24 patients reported pain of the knee and hip. On average, patients were 94.7 kg in weight. As shown in Table 2, there were no statistically significant differences between the treatment groups at baseline. There were also no differences in demographic and baseline characteristics (age, sex, duration of OA, perceived benefit of current analgesic, initial pain intensity) between patients who completed the study and those who did not.

All variables in this efficacy analysis favored controlled release codeine over placebo, as shown in Table 3. Using mean baseline values as the denominator, and mean values for the last week in the study (Week 4), percentage improvement

Table 1. Patient disposition.

	CR Codeine	Placebo	Total
Completion status			
Completed	31	35	66
Incomplete*	20	17	37
Total enrolled	51	52	103
Reason for non-completion			
Adverse event	15	4	19
Unrelated illness	1	0	1
Inadequate pain control	1	5	6
Patient noncompliant	1	1	2
Patient withdrawal	1	1	2
Protocol violation	0	1	1
Other reasons	1	5	6

Rate of non-completion by treatment, p = 0.54. Fisher's exact test.
CR: controlled release.

Table 2. Demographic and baseline data.

Characteristic	Completers CR Codeine, Mean (SD)	Completers Placebo, Mean (SD)	Non-Completers, Mean (SD)	Completers CR Codeine vs Placebo (p)
Age	60.1 (11.4)	63.0 (10.9)	63.1 (8.9)	0.29
Sex	18F:13M	23F:12M	23F:14M	0.61
Duration of OA (yrs)	11.8 (8.0)	9.5 (7.0)	9.9 (7.3)	0.22
Weight (kg)	94.5 (23.5)	94.8 (19.0)	94.6 (21.0)	0.96
Height (cm)	168.6 (10.5)	164.9 (10.3)	165.5 (10.7)	0.15
Knee OA	28	30	20 CC: 16 PL	0.71
Hip OA	13	19	5 CC: 12 PL	0.34
Baseline pain VAS (1–100 mm)	58.2 (18.9)	53.2 (24.5)	61.3 (19.8)	0.35
Pain at rest	24	27	29	
(% Yes)	77.4	77.1	76.1	0.8
Perceived benefit of current analgesic	1.7	1.8	1.6	
	(0.7)	(0.8)	(0.9)	0.84
Baseline WOMAC VAS pain scale	263.5	252.4	278.7	
(0–500 mm)	(99.7)	(129.8)	(133.2)	0.69
Baseline WOMAC VAS stiffness scale (0–200 mm)	126.5 (40.1)	106.2 (47.7)	131.4 (45.2)	0.07
Baseline WOMAC VAS function scale	900.5	844.9	867.7	
(0–1700 mm)	(357.3)	(405.3)	(398.4)	0.56
Baseline acetaminophen	9.5	11.1	11.2	
(tablets used)	(8.5)	(10.3)	(9.9)	0.51

CR: controlled release.

Table 3. Response characteristics (completers), mean (SD).

Outcome	Controlled Release Codeine, n = 31 Baseline	Week 4	Difference	Placebo, n = 35 Baseline	Week 4	Difference	p Codeine vs Placebo
WOMAC VAS pain (VAS, 0–500 mm)	263.5 (99.7)	145.4 (101.3)	118.0 (106.3)	252.4 (120.8)	221.3 (118.7)	31.1 (92.0)	0.0004
Weekly pain intensity (VAS, 0–100 mm)	65.4 (20.4)	29.4 (20.9)	36.0 (27.6)	57.4 (26.7)	47.8 (25.6)	8.9 (22.2)	0.0001
Pain over 24 h (VAS, 0–100 mm)	58.2 (18.9)	32.5 (21.4)	25.7 (23.3)	53.2 (24.5)	47.7 (24.7)	5.4 (20.3)	0.0001
WOMAC stiffness (VAS, 0–200 mm)	126.5 (40.1)	66.2 (46.3)	60.3 (51.1)	106.2 (47.7)	87.1 (52.8)	18.1 (40.7)	0.0030
WOMAC physical function (VAS, 0–1700 mm)	900.5 (357.3)	456.2 (316.2)	444.2 (400.8)	844.9 (405.3)	687.5 (415.5)	143.5 (284.7)	0.0007
Trouble falling asleep (VAS, 0–100 mm)	40.7 (37.2)	11.2 (21.2)	29.5 (37.5)	38.2 (34.5)	23.8 (25.5)	14.4 (34.7)	0.0220
Need medications to sleep (VAS, 0–100 mm)	34.5 (41.2)	9.3 (21.9)	25.3 (36.4)	24.9 (33.1)	22.3 (30.3)	2.6 (27.0)	0.0039
Pain on awakening (VAS, 0–100 mm)	36.8 (34.9)	21.5 (27.6)	28.1 (33.7)	39.5 (35.7)	30.9 (31.1)	9.0 (28.3)	0.0231
Rescue acetaminophen administration by week (Baseline is Week 1)	8.9 (7.2)	4.2 (5.8)	4.7	9.9 (8.0)	9.2 (8.1)	0.7	0.0051

was calculated for the major outcome variables. For the WOMAC pain VAS there was an improvement of 44.8% in the controlled release codeine group compared with 12.3% taking placebo ($p = 0.0004$). For the WOMAC stiffness scale the improvements were 47.7% and 17.0%, respectively ($p = 0.0030$). On the WOMAC physical function scale, the improvements were 49.3% and 17.0%, respectively ($p = 0.0007$).

Improvements in pain (vs placebo) seen on the WOMAC pain VAS increased with duration of treatment from week to week, starting at a mean improvement of 7.9 mm at Week 1, followed by 26.8 mm at Week 2, 53.6 mm at Week 3, and 75.9 mm at Week 4. Similar improvements in pain were also seen from the weekly VAS pain score, starting with an improvement of 3.3 mm at Week 1, 1.4 mm at Week 2, 16.2 mm at Week 3, and finally 18.4 mm at Week 4.

Figures 1 and 2 show the change from baseline at each week of the study for the WOMAC pain VAS and the weekly pain VAS score, respectively. There was a significant week by drug interaction for both the WOMAC pain VAS score ($p = 0.0195$) and the weekly pain score ($p = 0.0005$), reflecting the improvement in scores over the 4 weeks of the study with controlled release codeine, and lack of change over time with placebo. The differences between treatments for the WOMAC pain score and the weekly pain scale, respectively, were significant at Week 2 ($p = 0.0438$; 0.1677), Week 3 ($p = 0.0023$; 0.0001), and Week 4 ($p = 0.0004$; 0.0001), and overall ($p = 0.0030$; 0.0009). The mean doses of controlled release codeine in Weeks 1, 2, 3, and 4 of the study were 50.0 ± 0.0, 96.0 ± 13.1, 131.5 ± 38.2, and 158.9 ± 51.9 mg q12h, respectively, indicating a dose-response relationship for the effect of controlled release codeine.

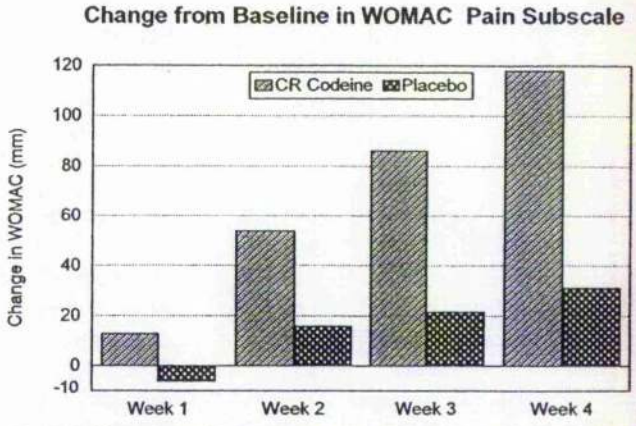


Figure 1. Mean change from baseline for the WOMAC Pain VAS scale at each week, with increasing dose from Weeks 1 to 4.

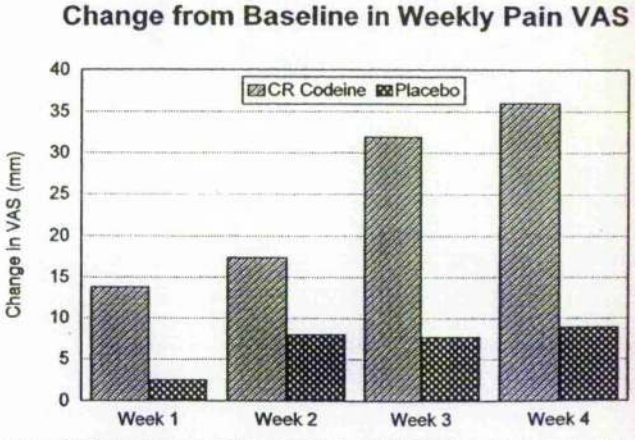


Figure 2. Change from baseline for the weekly VAS pain scale at each week, with increasing dose from Weeks 1 to 4.

Variables related to sleep all showed superiority of controlled release codeine over placebo, with less need for medication (improvements of 73.3% and 10.4%, respectively; p = 0.0039), less pain on awakening (improvements of 76.4% and 22.8%, respectively; p = 0.0231), and less trouble falling asleep (improvements of 72.5% and 37.7%, respectively; p = 0.0220).

The need for supplemental acetaminophen was also less in the controlled release codeine group (4.2 ± 5.8 rescue administrations/day) than during placebo treatment (9.2 ± 8.1 rescue administrations/day; p = 0.005).

Table 4 shows the global assessment scores. The patient clinical effectiveness evaluation was measured at study completion, and was rated at 2.1 ± 0.9 points (moderately effective) in the controlled release codeine group and 0.9 ± 1.0 (less than slightly effective) in the placebo group (p = 0.0001). Similarly, the investigators rated clinical effectiveness as being superior in the controlled release codeine group compared with placebo (1.9 ± 0.9 and 0.9 ± 1.0, respectively; p = 0.0001). The results of the Drug Liking Index showed a significant preference for codeine over placebo (p = 0.0011).

For all patients randomized to treatment, a significantly larger proportion (p < 0.01) of controlled release codeine patients experienced the following side effects than in the placebo group: constipation (49%, 11%), somnolence (39%, 10%), dizziness (33%, 8%), and overall (82%, 58%). For nausea, the proportions did not differ significantly (p = 0.091) by treatment group. Fifteen patients in the controlled release codeine group discontinued treatment due to adverse events compared with 4 patients treated with placebo. Among patients who completed the 4 week treatment period, only constipation, somnolence, and dizziness were found significantly more frequently on controlled release codeine (p < 0.05). The proportions of completed patients experiencing side effects overall were 80.7% and 62.9% for controlled release codeine and placebo, respectively (p = 0.173). The percentage of all completed controlled release codeine patients who experienced severe constipation, somnolence, dizziness, or nausea was 7.8%, 7.8%, 3.9%, and 0%, respectively. The percentage of all randomized controlled release codeine patients who experienced severe constipation, somnolence, dizziness, or nausea was 25.5%, 15.7%, 11.8%, and 3.9%, respectively.

Our objective was to investigate the efficacy of controlled release codeine and the primary analysis was limited to completed patients. To confirm the applicability of the results to patients who did not fully complete the study, 4 of the key outcome variables, WOMAC pain, daily VAS pain, WOMAC function, and WOMAC stiffness were subjected to an intent-to-treat analysis in which the mean scores from each patient's last week of treatment were used. The mean WOMAC pain scores for controlled release codeine (n = 48) and placebo (n = 46) were 164.2 ± 97.0 and 242.3 ± 112.2 (p = 0.0001); mean daily VAS pain scores were 37.5 ± 20.4 and 52.9 ± 24.9 (p = 0.0003). The mean WOMAC function scores for controlled release codeine (n = 47) and placebo (n = 46) were 551.3 ± 333.1 and 804.2 ± 448.2 (p = 0.0005) and mean WOMAC stiffness scores were 76.6 ± 48.6 and 101.7 ± 57.5 (p = 0.0073).

Table 4. Global assessments, mean (SD).

Outcome	CR Codeine, (n = 31)	Placebo, (n = 35)	p
Clinical effectiveness physician assessment	1.9 (0.9)	0.9 (1.0)	0.0001
Clinical effectiveness patient assessment	2.1 (0.9)	0.9 (1.0)	0.0001
Drug Liking Index	6.4 (2.0)	5.0 (1.6)	0.0011

CR: controlled release.

DISCUSSION

All the outcome variables in this study indicated the superiority of controlled release codeine over placebo in the treatment of pain due to OA of the hip and knee over a 4 week period. Although supplemental opioid containing preparations appear to be widely used clinically, in patients with OA, there are few objective studies demonstrating their effectiveness. For the WOMAC pain VAS scale, the primary outcome measure related to pain control, the decrease in score from baseline was 118 mm (45%) after 4 weeks' treatment. On the other dimensions of the WOMAC index, decreases of 66 mm (48%) and 444 mm (49%) were found for joint stiffness and physical function, respectively. These improvements in effect are at least as large as those seen after similar duration of treatment with the NSAID meclofenamate[40], tenoxicam[43], and diclofenac[40,43]. The trial duration was shorter than that often used in the evaluation of NSAID and longer term controlled trials will be needed. Recent data from a retrospective study[44] of a cohort of more than 600 rheumatology clinic patients suggests that longterm benefit of opioid treatment may be expected in controlled trials. Ytterberg, *et al*[44] found 21% of clinic patients had used opioids for more than 3 months and 24% had used opioids for less than 3 months. In both groups, patients reported significant pain relief that had been maintained for up to 3 years.

A possible criticism of our study is that patients were not asked to guess whether they had received active treatment or placebo. Both the experience of prominent opioid side effects and the relief of pain represent a way in which the identity of the assigned treatment can become evident to patients in analgesic trials, particularly those involving opioids.

In our study, patients were withdrawn from NSAID prior to initiation of study drugs. However, recent controlled studies have also shown that opioids can produce further reduction in pain even in patients treated with maximal doses of NSAID. In a double blind placebo controlled study, controlled release oxycodone at a dose of 20 mg q12h produced a 36% decrease in pain intensity, measured on an ordinal scale, in a population

in which a majority of the patients were maintained on their prestudy dose of NSAID[45]. Similarly, in another study in which all patients received optimal doses of NSAID the addition of oxycodone resulted in a 43% reduction in ordinal pain score[46] and the reduction in pain achieved with controlled release oxycodone was the same as that achieved with the same daily dose of oxycodone administered in combination with acetaminophen.

Although both codeine and oxycodone are usually administered in fixed dose combination with acetaminophen or ASA, results such as these question the value of using such combinations in patients already treated with NSAID. Several studies in postoperative pain have documented that the analgesia produced by non-opioid/opioid combinations is essentially equivalent to the added effects of the components[47]. Similarly, in patients with OA awaiting hip joint surgery the addition of 30 mg codeine to 200 mg ibuprofen provided significantly greater analgesia[48]. It is therefore likely that patients already treated with maximal doses of NSAID will not realize any benefit from added acetaminophen or ASA in combination with codeine or oxycodone and may experience greater risk of toxicity. In addition, single entity opioid formulations have the advantage of availability of 12 hourly controlled release formulations and no restriction on the ability to titrate the dose to optimal effect due to the maximum recommended doses of acetaminophen or ASA.

In addition to a beneficial effect of controlled release codeine in reducing OA pain, patients in this study also reported improvements in quality of sleep and physical function, important determinants of quality of life in patients with chronic musculoskeletal disease. Since lack of adequate sleep can exacerbate pain, improvement of sleep quality is an important goal of pain management[49]. The improvement in sleep quality found in this study may be attributable to the prolonged duration of action of the controlled release formulation as evidenced by the lower level of pain on awakening. In addition, the reduced use of medication to get to sleep indicates that the better pain control achievable with around-the-clock dosing may have facilitated the onset of sleep. Previous studies have shown a close association between measurements of pain and disability[37]. The finding of a similar degree of improvement in both the pain and physical function WOMAC scales in the controlled release codeine group is therefore not surprising. Similar improvements in physical function, assessed using the Pain Disability Index, were also found in a placebo controlled trial of controlled release codeine in patients with chronic nonmalignant pain of mixed etiology[20]. These observations confirm the important effect that adequate relief of pain can have on quality of life of patients with OA.

There were also significant improvements in joint stiffness in the patients treated with controlled release codeine. While a number of factors, such as muscle tone and conditioning, contribute to joint stiffness, its relationship to joint inflammation raises the possibility that the effect of codeine on joint stiffness may involve a peripheral opioid mechanism. Opioid receptors located on peripheral terminals of primary afferent sensory neurones are upregulated in inflamed tissue, and the action of opioid agonists on these receptors inhibits the release of excitatory proinflammatory mediators[50,51]. Inflammatory cells such as lymphocytes, macrophages, and mast cells contain endogenous opioid peptides that may play a role in modulating both hyperalgesia and inflammation[52]. Inhibition of neuropeptide release represents a peripheral mechanism by which opioids could reduce inflammation induced hyperalgesia and provide a possible explanation for the direct antiinflammatory effect of opioids observed in animal models of arthritis[53]. These observations, together with the demonstration of an effect of codeine on both pain and joint stiffness in patients with OA, suggest that additional studies of opioids are needed in conditions, such as rheumatoid arthritis, with a more prominent inflammatory component and hyperalgesia in the form of joint tenderness.

A major consideration in the evaluation of opioid analgesic therapy for chronic nonmalignant pain is the perception that patients with chronic pain receiving opioid analgesics are at high risk of iatrogenic addiction. This can have significant influence on clinicians' willingness to treat chronic nonmalignant pain with opioids because of failure to distinguish physical dependence from psychological dependence. Physical dependence is characterized by the occurrence of an abstinence syndrome following abrupt cessation of an opioid agonist or following the administration of an opioid antagonist. Addiction, on the other hand, is a psychological and behavioral syndrome characterized by an intense desire for the opioid, along with evidence of compulsive drug use and acquisition of opioids by manipulation of the medical system or from a nonmedical source[15]. Except in individuals who have a history of substance abuse, addiction is not a common observation in patients who take opioids to control pain[15]. The potential for abuse was assessed in this study using a Drug Liking Index[39], which showed a slightly higher score on codeine compared with placebo, although we did not observe unusual requests for additional codeine. Although patients were instructed to rate their liking for the drug based only on its effects other than analgesia, there is a possibility that some patients' assessments are influenced by the pain relief they receive. In a crossover study requiring treatment for 9 weeks with morphine or placebo[19], there was a nonsignificant but higher Drug Liking Index score with placebo, suggesting that the risk of addiction in patients with chronic nonmalignant pain treated with opioids is very low.

This double blind randomized controlled study demonstrates the superiority of controlled release codeine over placebo in the treatment of pain due to OA of the hip or knee. This benefit was observed across all outcome measures including improvements in joint stiffness, quality of sleep, and physical function, as well as measures of OA pain. In

addition, there was a progressive increase in efficacy as the dose of controlled release codeine was increased at weekly intervals over the 4 weeks of the study. A high proportion of patients in both treatment groups reported typical opioid side effects such as constipation, somnolence, dizziness, and nausea. It is possible that a more gradual dose escalation would have resulted in fewer patients discontinuing codeine due to adverse events.

ACKNOWLEDGMENT

The authors are grateful to Iva Ford for study monitoring and for comments on the manuscript.

REFERENCES

1. Bellamy N, Bradley L. Workshop on chronic pain, pain control, and patient outcomes in rheumatoid arthritis and osteoarthritis. Arthritis Rheum 1996;39:357-62.
2. Hochberg M, Altman R, Brandt K, et al. Guidelines for the medical management of osteoarthritis. Arthritis Rheum 1995;38:1535-46.
3. Tannenbaum H, Davis P, Russell A, et al. An evidence-based approach to prescribing NSAID's in musculoskeletal disease: A Canadian consensus [review]. Can Med Assoc J 1996;155:77-88.
4. Peloso P. Degenerative joint disease. Therapeutic Choices. Ottawa: Canadian Pharmaceutical Association; 1995:338-43.
5. Bradley J, Brandt KD, Katz BP, Kalasinski LA, Ryan SI. Comparison of an anti-inflammatory dose of ibuprofen, an analgesic dose of ibuprofen and acetaminophen in the treatment of patients with osteoarthritis of the knee. N Engl J Med 1991; 325:87-91.
6. Wilcox CM, Shalek KA, Cotsonis G. Striking prevalence of over-the-counter nonsteroidal antiinflammatory drug use in patients with upper gastrointestinal hemorrhage. Arch Intern Med 1994;154:42-5.
7. Guthann S, Rodriguez G, Raiford D. Individual nonsteroidal antiinflammatory drugs and other risk factors for upper gastrointestinal bleeding and perforation. Epidemiology 1997; 8:18-24.
8. Soll AH, Weinstein WM, Kurata J, McCarthy D. Nonsteroidal antiinflammatory drugs and peptic ulcer disease. Ann Intern Med 1991;114:307-19.
9. Clive DM, Stoff JS. Renal syndromes associated with nonsteroidal antiinflammatory drugs. N Engl J Med 1984;310:563-72.
10. Perneger TV, Whelton PK, Klag MJ. Risk of kidney failure associated with the use of acetaminophen, aspirin, and nonsteroidal antiinflammatory drugs. N Engl J Med 1994;331:1675-9.
11. Segasothy M, Suleiman AB, Puvaneswary M, Rohanna A. Paracetamol: a cause for analgesic nephropathy and end-stage renal disease. Nephron 1988;50:50-4.
12. Sandler DP, Smith JC, Weinberg CR, et al. Analgesic use of chronic renal disease. N Engl J Med 1989;320:1238-43.
13. Jacox AK, Carr DB, Payne R, et al. Management of cancer pain, clinical practice guideline. 9. Rockville, MD: Agency for Health Care Policy and Research; 1994 (AHCPR Publ. No. 94-D592).
14. Levy MH. Pharmacologic treatment of cancer pain. New Engl J Med 1996;335:1124-32.
15. Portenoy R. Opioid therapy for chronic nonmalignant pain: current status. Pain Res Manage 1994;1:247-87.
16. Zenz M, Strumpf M, Tryba M. Long-term oral opioid therapy in patients with chronic non-malignant pain. J Pain Symptom Manage 1992;7:69-77.
17. Schofferman J. Long-term use of opioid analgesics for the treatment of chronic pain of nonmalignant origin. J Pain Symptom Manage 1993;7:279-88.
18. France RD, Urban BJ, Keefe FJ. Long term use of narcotic analgesics in chronic pain. Soc Sci Med 1984;19:1379-82.
19. Moulin DE, Iezzi A, Amireh R, Sharpe WK, Boyd D, Merskey H. Randomised trial of oral morphine for chronic non-cancer pain. Lancet 1996;347:143-7.
20. Arkinstall W, Sandler A, Goughnour B, Babul N, Harsanyi Z, Darke A. Efficacy of controlled release codeine in chronic non-malignant pain: a randomized, placebo controlled clinical trial. Pain 1995;62:169-78.
21. Hale ME, Speigth KL, Harsanyi Z, et al. Efficacy of 12 hourly controlled-release codeine compared with as required dosing of acetaminophen plus codeine in patients with chronic low back pain. Pa Res Manage 1997;2:33-8.
22. Anonymous. Guidelines for the management of chronic non-malignant pain, February 1993. Bulletin of the College of Physicians and Surgeons of Alberta, 15 March 1993, Edmonton, Alberta.
23. Goldman B. A physician's guide to chronic opioid therapy. The Can J CME 1997;9:71-9.
24. Dieppe P. Therapeutic targets in osteoarthritis. J Rheumatol 1995;43:136-9.
25. Muskowitz RW. Clinical and laboratory findings in osteoarthritis. In: McCarty DJ, editor. Arthritis and allied conditions. A textbook of rheumatology. 11th ed. Philadelphia: Lea and Febiger; 1989: Ch.102.
26. Stein C, Comisel K, Haimerl E, et al. Analgesic effect of intraarticular morphine after arthroscopic knee surgery. N Engl J Med 1991;325:1123-6.
27. Khoury GF, Chen AC, Geriand DE, Stein C. Intraarticular morphine, bupivacaine, and morphine/bupivacaine for pain control after knee videoarthroscopy. Anesthesiology 1992;77:263-6.
28. Joshi GP, McCarroll SM, O'Brien TM, Lenane P. Intra-articular analgesia following knee arthroscopy. Anesth Analg 1993;76:333-6.
29. De Andres J, Beliver J, Barrera L, Febre E, Bolinches R. A comparative study of analgesia after knee surgery with intraarticular bupivacaine, intraarticular morphine, and lumbar plexus block. Anesth Analg 1993;77:727-30.
30. Haynes TK, Appadurai IR, Power I, Rosen M, Grant A. Intraarticular morphine and bupivacaine analgesia after arthroscopic knee surgery. Anesthesia 1994;49:54-6.
31. Heine MF, Tillel ED, Tsueda K, et al. Intra-articular morphine after arthroscopic knee operations. Br J Anaesthesiol 1994;73:413-5.
32. Dalsgaard J, Felsby S, Juelsgaard P, Frockjaer J. Low-dose intra-articular morphine analgesia in day care knee arthroscopy: a randomized double-blinded prospective study. Pain 1994;56:151-4.
33. Ferrel B, Wisdom C, Wenzl C, Bown J. Effects of controlled-release morphine on quality of life for cancer pain. Oncol Nurs Forum 1989;16:521-6.
34. Band CJ, Band PR, Deschamps M, Besner J, Goldman A. Human pharmacokinetic study of immediate release (codeine phosphate) and sustained-release (Codeine Contin) codeine. J Clin Pharmacol 1994;34:938-43.
35. Chary S, Goughnour BR, Moulin DE, Thorpe WR, Harsanyi Z, Darke AC. The dose-response relationship of controlled-release codeine (Codeine Contin) in chronic cancer pain. J Pain Symptom Manage 1994;9:363-71.
36. Atlas of standard radiographs. The epidemiology of chronic rheumatism. Vol 2. Philadelphia: F.A. Davis; 1963.
37. Bellamy N, Buchanan W, Goldsmith C, Campbell J, Stilt L. Validation study of WOMAC: A health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988;15:1833-49.
38. Bellamy N. Pain assessment in osteoarthritis. Experience with the WOMAC osteoarthritis index. Semin Arthritis Rheum 1989; 18:14-7.

39. Jasinski DR. Assessment of the abuse potentiality of morphine-like drugs (methods used in man). In: Martin WR, editor. Drug addiction. I. Morphine, sedative hypnotic and alcohol dependence. Handb Exp Pharmacol 1977;45:197-258.
40. Bellamy N, Carette S, Ford PM, et al. Osteoarthritis antirheumatic drug trials. III. Setting the delta for clinical trials — results of a consensus development (Delphi) exercise. J Rheumatol 1992;19:451-7.
41. Bellamy N, Kean WF, Buchanan WW, Gerecz-Simon E, Campbell J. Double-blind randomized controlled trial of sodium meclofenamate (Meclomen) and diclofenac sodium (Voltaren); Post-validation reapplication of the WOMAC Osteoarthritis Index. J Rheumatol 1992;19:153-9.
42. Bellamy N, Carette S, Ford PM, et al. Osteoarthritis antirheumatic drug trials. II. Tables for calculating sample size for clinical trials. J Rheumatol 1992;19:444-50.
43. Bellamy N, Buchanan WW, Chalmers A, et al. A multicentre study of tenoxicam and diclofenac in patients with osteoarthritis of the knee. J Rheumatol 1993;20:999-1004.
44. Ytterberg SR, Mahowald ML, Woods SR. Codeine and oxycodone use in patients with chronic rheumatic disease pain. Arthritis Rheum 1998;41:1603-12.
45. Lacouture P, Roth S, Burch F, Fleischmann R, Iwan T, Kaiko R. A double-blind, placebo-controlled, repeated dose, dose-response evaluation of controlled-release (CR) oxycodone in patients with osteoarthritis. Clin Pharmacol Therapeutics 1996;59:129.
46. Caldwell JR, Hale ME, Boyd RE, et al. Treatment of osteoarthritis pain with controlled-release oxycodone or fixed-combination oxycodone plus acetaminophen added to nonsteroidal antiinflammatory drugs: A double-blind, randomized, multicenter, placebo-controlled trial. J Rheumatol 1999;26:862-9.
47. Gertzbein SD, Tile M, McMurty RY, et al. Analysis of the analgesic efficacy of acetaminophen 100 mg, codeine phosphate 60 mg, and the combination of acetaminophen 100 mg and codeine phosphate 60 mg in the relief of postoperative pain. Pharmacotherapy 1986;6:104-7.
48. Quiding H, Grimstad J, Rusten K, Stubhaug A, Bremnes J, Breivik H. Ibuprofen plus codeine, ibuprofen, and placebo in a single dose and multidose cross-over comparison for coxarthrosis pain. Pain 1992;50:303-7.
49. Holman GH. Chronic nonmalignant pain. Clin Geriatr 1997; 5:21-41.
50. Stein C. The control of pain in peripheral tissues by opioids. New Engl J Med 1995;332:1685-90.
51. Hargreaves KM, Joris JL. The peripheral analgesic effects of opioids. Am Pain Soc J 1993;2:51-9.
52. Stein C, Hassan AHS, Lehrberger K, Giefing J, Yassouridis A. Local analgesic effect of endogenous opioid peptides. Lancet 1993;342:321-4.
53. Barber A, Gottschlich R. Opioid agonists and antagonists: an evaluation of their peripheral actions in inflammation. Med Res Rev 1992;12:525-62.

ARTHRITIS & RHEUMATISM
Vol. 43, No. 10, October 2000, pp 2339–2348

EFFICACY AND SAFETY OF DIACEREIN IN OSTEOARTHRITIS OF THE KNEE

A Double-Blind, Placebo-Controlled Trial

JEAN-PIERRE PELLETIER, MICHAEL YARON, BOULOS HARAOUI, PATRICK COHEN, MENAHEM A. NAHIR, DENIS CHOQUETTE, IRINA WIGLER, ITZHAK A. ROSNER, ANDRÉ D. BEAULIEU, and The DIACEREIN STUDY GROUP

Objective. To evaluate the efficacy and safety of diacerein, a drug with interleukin-1β–inhibitory activity in vitro, in patients with knee osteoarthritis (OA).

Methods. A total of 484 patients fulfilling the American College of Rheumatology criteria for knee OA were enrolled in this 16-week, randomized, double-blind, placebo-controlled, parallel study group with 3 diacerein dosages of 50 mg/day, 100 mg/day, and 150 mg/day (administered twice daily).

Results. In the intent-to-treat population, 100 mg/day diacerein (50 mg twice daily) was significantly superior ($P < 0.05$) to placebo using the primary criterion (visual analog scale [VAS] assessment of pain on movement). Significant improvement ($P < 0.05$) was also observed for the secondary criteria, which included the Western Ontario and McMaster Universities OA Index (WOMAC), the WOMAC subscores, and the VAS assessment of handicap. In patients treated with diacerein dosages of 50 mg/day and 150 mg/day, favorable but not significant results were observed for the primary criterion. The best daily dosage of diacerein, calculated from the effect on the VAS assessment of pain on movement, was 90.1 mg. In the per-protocol population, the analysis of the primary criterion showed significant dose-dependent differences ($P < 0.05$) between each of the 3 diacerein dosages and the placebo. No differences were observed among the 3 diacerein groups. A significantly higher incidence ($P < 0.05$) of adverse events (AEs), as well as a higher rate of dropoout due to AEs, was observed in patients treated with 150 mg/day diacerein versus those treated with placebo, 50 mg/day diacerein, or 100 mg/day diacerein. Mild-to-moderate transient changes in bowel habits were the most frequent AEs, increasing with the dosage.

Conclusion. Diacerein, a drug for the treatment of OA, was shown to be an effective treatment for symptoms in patients with knee OA. Taking into account both efficacy and safety, the optimal daily dosage of diacerein for patients with knee OA is 100 mg/day (50 mg twice daily).

Diacerein is a drug with interleukin-1 (IL-1)–inhibitory activity developed for the treatment of osteoarthritis (OA) (1). In animals, oral administration of diacerein resulted in antiinflammatory activity as manifested by an inhibition of edema induced by the injection of carrageenan into the footpad (2). Diacerein inhibited adjuvant arthritis induced in rats by the injection of *Mycobacterium tuberculosis* (1). It also exhibited analgesic effects and antipyretic activities in animal models (1). Diacerein has no inhibitory effect on phospholipase A_2, cyclooxygenase (COX), or lipooxygenase in vitro; on the contrary, diacerein stimulates prostaglandin E_2 synthesis in human chondrocyte cultures (3). Both diacerein and its active metabolite rhein are powerful inhibitors in vitro of the synthesis of cytokines (3–5), mainly IL-1β,

Supported by a grant from Les Laboratoires Negma, Toussus-le-Noble, France.

Jean-Pierre Pelletier, MD, Boulos Haraoui, MD, Denis Choquette, MD: Centre hospitalier de l'Université de Montréal, Hôpital Notre-Dame, Montreal, Quebec, Canada; Michael Yaron, MD, Irina Wigler, MD: Ichilov Hospital, Souraski Medical Center, Tel Aviv, Israel; Patrick Cohen, MD: Laboratoires Negma, Toussus-le-Noble, France; Menahem A. Nahir, MD: Rambam Medical Center, Haifa, Israel; Itzhak A. Rosner, MD: Bnei Zion Medical Center, Haifa, Israel; André D. Beaulieu, MD: Centre hospitalier de l'Université Laval, Quebec City, Quebec, Canada.

Address reprint requests to Jean-Pierre Pelletier, MD, Unité de recherche en Arthrose, Centre hospitalier de l'Université de Montréal, Hôpital Notre-Dame, 1560 rue Sherbrooke Est, Montreal, Quebec City H2L 4M1, Canada.

Submitted for publication March 2, 2000; accepted in revised form June 27, 2000.

and metalloproteases (collagenase and stromelysin) (6,7) that are involved in cartilage degradation. The IL-1β-inhibitory activity of diacerein has been shown in vitro to be related to the inhibition of cytokine synthesis per se in synovium and chondrocytes, as well as to a reduction in the level of IL-1 receptor bioactivity (4,5) in these cells. The capacity of diacerein to reduce the structural changes of OA has been shown in several in vivo experimental models of surgically or mechanically induced OA, such as the dog, sheep, guinea pig, and rabbit models (8–11).

Several double-blind, randomized, controlled clinical trials (12–17) have demonstrated a beneficial effect of diacerein on the clinical signs or symptoms of OA, comparable with that of nonsteroidal antiinflammatory drugs (NSAIDs). Diacerein is a slow-acting agent with symptomatic effects appearing 4 weeks after beginning treatment; most interesting are the results of a study conducted in patients with hip OA which indicated that, in these patients, diacerein's effect was additive with that of NSAIDs (12). In several studies, sustained pain relief was observed for several weeks after discontinuation of diacerein, suggesting the presence of a carryover effect of the drug (13,15,17).

According to previous studies of the drug's profile, the risk:benefit ratio of diacerein for the OA patient is better than that of NSAIDs (18). This also seems to be the case with respect to the administration of diacerein to elderly patients who frequently have multiple illnesses and are predisposed to gastrointestinal (GI) tract complications due to long-term administration of NSAIDs (19).

The purpose of this double-blind, placebo-controlled, phase II dose-range study was to assess the optimal daily dosage of diacerein to be given to patients with knee OA.

PATIENTS AND METHODS

Patients. Outpatients of either sex, ages 40–80 years, with tibiofemoral OA (grades I–III of the Kellgren/Lawrence [K/L] classification [20]), fulfilling the American College of Rheumatology criteria for knee OA (21), and with pain present most days of the prior month were recruited for the study. Radiographic evidence of knee OA was defined by the presence of osteophytes in at least 1 tibiofemoral compartment (the radiographs having been obtained <6 months prior to enrollment and with at least 2 views, posteroanterior and lateral). For study enrollment, there had to be evidence of knee pain on movement scored by the patients at ≥35 mm on a 100-mm visual analog scale (VAS). Patients were not retained for the study if they had serious concomitant medical illness, secondary OA, radiographic grade IV by the K/L classification, or knee surgery planned within the following 6 months. Patients were not to have been treated with any drug supposed to be structure modifying in OA, nor were they to have been treated with any intraarticular injection of corticosteroids for at least 3 months before the study.

Before entering the trial, patients underwent washout periods of 7 days for any NSAIDs or 12 hours for analgesics. During the trial, acetaminophen intake (500-mg tablets) was permitted in cases of persistent pain, and the dose and duration were recorded. Patients were enrolled in 15 centers in Canada and 10 centers in Israel. Two committed clinical research organizations, one in Israel (More Research Application) and the other in Canada (Integrated Research Incorporation), were in charge of overseeing the study, which included verifying both the investigators' professional qualifications and the meeting of selection criteria by patients, as well as monitoring the 7 scheduled visits.

Study design. This study was a prospective, multicenter, randomized, double-blind, placebo-controlled, parallel, 4-arm trial of 16-week duration. This duration was chosen according to the results observed in several studies, which demonstrated a delay of action by diacerein on the signs and symptoms of OA (≥45 days) (12–15), and also to gain more information about diacerein's safety profile.

The study was conducted in accordance with the Helsinki Declaration (1964) and its revision (1975), and was approved by the institutional review boards of all Israeli and Canadian study sites. Patients entered the study after fulfilling the inclusion and exclusion criteria and signing an informed-consent form.

Drug administration. Patients were randomly assigned to 4 treatment groups. The centralized allocation schedule was prepared using a blocked randomization technique (blocking factor 8). The treatments were divided between the 2 countries (treatments 1–500 in Israel and 600–1,000 in Canada) and then allocated to each center. One group received placebo (1 capsule twice daily), the second group received 50 mg diacerein (25 mg twice daily), the third group received 100 mg diacerein (50 mg twice daily), and the last group received 150 mg diacerein (75 mg twice daily).

Evaluation of efficacy. The primary efficacy parameter was the patient's assessment of pain on movement (for the 48 hours prior to the visit) using a 100-mm VAS ranging from 0 = no pain to 100 = unbearable pain. The main secondary efficacy criteria included the following: 1) the Western Ontario and McMaster Universities OA Index (WOMAC) version VA 3.0 (22,23) (for Israeli patients, the WOMAC was translated into Hebrew and the translation was validated [24]); 2) the VAS of handicap (100-mm VAS ranging from 0 = no impairment to 100 = unbearable impairment) assessed by asking the patient "What is your main problem associated with your osteoarthritis?" This scale was shown to be correlated with other functional disability scales in rheumatology (25,26) and was used in a previous clinical trial with diacerein (13); and 3) the patient and physician overall assessments expressed at the end of the treatment period on a 100-mm VAS (efficacy, ranging from 0 = very poor to 100 = excellent). The other secondary efficacy criteria included the following: 1) knee joint swelling according to a 0–3-grade scale (0 = absent, 1 = detectable without loss of bone contour, 2 = loss of bone contour, and 3 = important synovial thickening or synovial effusion); 2) duration

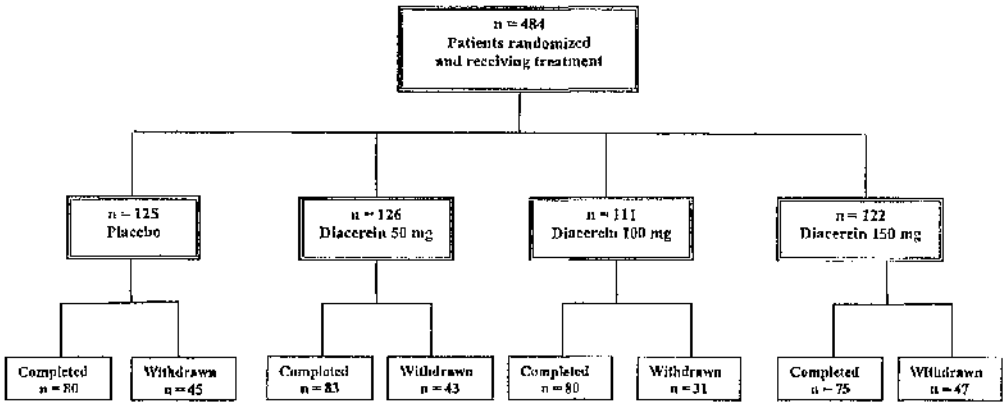


Figure 1. Flow chart of the disposition of the patients in the study.

of morning stiffness, in minutes; and 3) joint mobility measured by flexion and extension in degrees, assessed with a goniometer. All variables were assessed at baseline and at weeks 2, 4, 8, 12, and 16, or at the time of premature withdrawal.

Evaluation of safety. All adverse events (AEs) reported by the patients during the study treatment were recorded on the Case Report Form (CRF) and classified in terms of type, time of onset, severity (mild, moderate, or severe), duration, and outcome. The physician asked the patient "How did you tolerate the test medication?" and recorded the response. All the information concerning expected AEs was described on the informed-consent form. The physician was asked to express an opinion regarding the relationship of the AE to the study treatments.

Laboratory tests were performed at baseline and at weeks 4 and 16 (or at the time of premature withdrawal), including blood cell count, erythrocyte sedimentation rate, serum chemistries (electrolytes, liver enzymes, total bilirubin, uric acid, and creatinine), and urinalysis (using reagent sticks).

The global tolerance to the study treatment was assessed by the patient and the investigator at weeks 2–16 with a scoring system using a 5-point scale ("very good," "good," "moderate," "bad," and "very bad").

Compliance. Medication compliance and concomitant treatment such as acetaminophen were recorded on the CRF. Capsule counts were performed at each visit (weeks 2–16).

Statistical analysis. Sample size estimates, performed to detect a pre–post difference in VAS assessment of pain on movement of 10 mm, indicated that 100 patients were required for each treatment group with the level of significance (1-tailed test) set at $P < 0.05$ and a power of 90%. The comparisons between groups allowed the use of the 1-tailed statistical tests because of the constant superiority of diacerein treatment over placebo shown in previous studies (12–14). With an anticipated dropout rate of 20%, the sample size was increased to 125 patients per arm.

Three populations were analyzed in the study, 1 for safety and 2 for efficacy. The safety population was defined as all patients who were randomly assigned and who received study medication at least once. The 2 efficacy populations were the intent-to-treat population (ITT), which consisted of all patients who were part of the safety population and for whom we had at least 1 postbaseline measure of the primary efficacy criterion (VAS assessment of pain on movement); and the per-protocol population (PP), which consisted of all patients who completed the 16-week study.

All demographic and assessment variables were compared at baseline using an analysis of variance (2-tailed at the 5% significance level) for continuous variables and chi-square tests for categorical variables.

The primary efficacy criterion, VAS assessment of pain on movement, was first analyzed using the normalized area under the curve (AUC) for the difference from baseline scores (scores at baseline visit) from baseline to week 16 using the trapezoidal rule (27). A secondary analysis for this criterion was based on the average (AVE) of the score differences from baseline at each visit.

The last observation carried forward (LOCF) method and the linear interpolation method were used to estimate missing values. The main analysis was based on comparison of the placebo group with each of the treatment groups, using the Dunnett 1-tailed test to detect significant differences between the treatment groups and the placebo group. When significant differences were detected, a quadratic polynomial (28) was fitted and used to determine the best dosage (i.e., the dosage value that optimizes the quadratic function).

The main analysis of the secondary parameters was based on the Dunnett 1-tailed test to detect significant differences between the treatment groups and the placebo group. The AUC and AVE were used as response variables for all criteria except global efficacy assessment, which used raw score assessments at each visit.

All analyses were performed in SAS 6.12 (SAS Institute, Cary, NC) under HP-UX version 10.20 (Hewlett-Packard, McMinnville, OR). The level of significance was set at $P < 0.05$ (1-tailed test) for comparison with placebo and $P < 0.05$ (2-tailed test) for pairwise comparisons resulting from the overall treatment analyses. There were no adjustments for multiple comparisons other than the implicit adjustments used by the Dunnett test.

Table 1. Patient demographics and baseline characteristics*

	Placebo (n = 124)	50 mg/day diacerein (n = 126)	100 mg/day diacerein (n = 110)	150 mg/day diacerein (n = 120)
Age, years	64.5 ± 8.65	62.95 ± 8.41	64.22 ± 8.02	62.27 ± 10.18
Men, %	21.0	16.7	24.5	20.0
Women, %	79.0	83.3	75.5	80.0
Height, cm	160.94 ± 9.24	159.48 ± 8.66	161.54 ± 8.42	160.71 ± 7.58
Weight, kg	80.45 ± 14.48	80.55 ± 15.56	82.78 ± 17.52	80.62 ± 16.61
BMI, kg/m^2	31.05 ± 5.35	31.63 ± 5.50	31.73 ± 6.21	30.99 ± 5.88
Disease duration	8.0 ± 7.41	7.8 ± 7.18	8.1 ± 6.42	7.8 ± 6.99
Pain, VAS, mm	70.54 ± 19.05	67.27 ± 17.69	73.56 ± 16.74	69.85 ± 18.77
WOMAC, mm	180.79 ± 61.68	168.72 ± 59.48	181.92 ± 54.45	175.72 ± 59.66
Handicap, VAS, mm†	69.48 ± 20.52	65.73 ± 21.51	73.29 ± 17.41	68.22 ± 20.09
Joint mobility, degrees				
Extension	4.33 ± 5.08	4.45 ± 5.66	4.79 ± 5.19	4.98 ± 5.74
Flexion	110.24 ± 21.95	110.40 ± 24.76	109.45 ± 23.05	109.28 ± 23.30
Morning stiffness, minutes	21.48 ± 21.91	24.01 ± 50.60	17.15 ± 19.11	17.58 ± 19.71

* Except where otherwise indicated, values are the mean ± SD. BMI = body mass index; VAS = visual analog scale; WOMAC = Western Ontario and McMaster Universities Osteoarthritis Index.
† $P = 0.0353$ by analysis of variance between the groups.

RESULTS

Patient characteristics. A total of 484 patients enrolled in the study (Figure 1) by 42 Israeli and Canadian investigators were randomly assigned as follows: 125 (25.8%) received the placebo, 126 (26.0%) received 50 mg/day diacerein, 111 (22.9%) received 100 mg/day diacerein, and 122 (25.2%) received 150 mg/day diacerein.

The patients were primarily white (92.4%) women (79.6%), with a mean ± SD age of 63.5 ± 8.9 years. The mean ± SD body mass index was 31.3 ± 5.7 kg/m^2, and the mean ± SD time from diagnosis of OA (disease duration) was 7.9 ± 7.0 years. No major differences were observed among the 4 groups at baseline, except for VAS assessment of handicap (the largest observed difference was between the 50 mg/day and 100 mg/day diacerein groups) (Table 1).

Of the 484 randomly assigned patients, 4 (1 in the placebo group, 1 in the 100 mg/day diacerein group, and 2 in the 150 mg/day diacerein group) did not have postbaseline VAS assessment of pain on movement, and were therefore excluded from the efficacy analyses.

A total of 166 patients (34.3%) withdrew from the study: 36.0%, 34.1%, 27.9%, and 38.5% in the placebo, 50 mg/day diacerein, 100 mg/day diacerein, and 150 mg/day diacerein groups, respectively (Figure 1). The reasons for discontinuation were mainly the lack of efficacy of the study treatment for the patients receiving the placebo or 50 mg/day diacerein and the occurrence of AEs for patients receiving 150 mg/day diacerein (Table 2).

A total of 318 patients (65.7%) completed the study: 80 of 125 (64.0%) in the placebo group, 83 of 126 (65.9%) in the 50 mg/day diacerein group, 80 of 111 (72.1%) in the 100 mg/day diacerein group, and 75 of 122 (61.5%) in the 150 mg/day diacerein group (Figure

Table 2. Reasons for discontinuation of treatment*

Treatment	No. of randomly assigned patients	Reason for treatment discontinuation: Adverse events	Lack of efficacy	Other	No. of patients completing study
Placebo	125	14 (11.2)	23 (18.4)	8 (6.4)	80 (64.0)
Diacerein					
50 mg/day	126	16 (12.7)	21 (16.7)	6 (4.8)	83 (65.9)
100 mg/day	111	11 (9.9)	12 (10.8)	8 (7.2)	80 (72.1)
150 mg/day	122	23 (18.9)	16 (13.1)	8 (6.6)	75 (61.5)
Total	484	64 (13.2)	72 (14.9)	30 (6.2)	318 (65.7)

* Values are the number (%).

Table 3. Differences in clinical assessment criteria from baseline to week 24 in each of the groups in the intent-to-treat population*

	Placebo (n = 124)	50 mg/day diacerein (n = 126)	100 mg/day diacerein (n = 110)	150 mg/day diacerein (n = 120)
Pain, VAS, mm	−10.9 ± 19.3	−15.6 ± 21.0	−18.3 ± 19.3†	−14.3 ± 23.7
WOMAC, mm	−16.7 ± 51.9	−27.4 ± 52.7	−36.7 ± 52.3†	−29.1 ± 46.7
Pain	−33.9 ± 90.5	−50.3 ± 94.8	−58.8 ± 92.5†	−50.5 ± 88.1
Stiffness	−10.3 ± 42.4	−17.7 ± 43.4	−27.3 ± 42.3†	−21.4 ± 37.0
Physical	−85.8 ± 304.4	−139.3 ± 301.4	−193.3 ± 318.0†	−143.3 ± 278.8
Handicap, VAS, mm	−9.9 ± 20.1	−12.3 ± 22.7	−18.5 ± 22.1†	−12.8 ± 21.7
Global efficacy assessment, mm				
Investigator	41.4 ± 32.4	−48.6 ± 33.6	−48.8 ± 29.7	48.1 ± 32.7
Patient	42.4 ± 33.1	−49.2 ± 33.6	−51.4 ± 29.3	50.0 ± 32.3

* Values are the mean ± SD. See Table 1 for definitions.
† $P < 0.05$ versus placebo group.

1). Therefore, for the efficacy analyses, the ITT population consisted of 480 patients and the PP population consisted of 318 patients. All 484 randomly assigned patients were involved in the safety analysis.

Efficacy. The results of the analysis of the primary efficacy criterion, VAS assessment of pain on movement, showed the diacerein treatment to be superior to the placebo, irrespective of the diacerein dosage and of the population analyzed (Tables 3 and 4). For this criterion, a statistically significant difference ($P < 0.05$) was shown between the 100 mg/day diacerein and placebo groups for the ITT population (Table 3 and Figure 2). There were also statistically significant differences for the secondary criteria (mainly the WOMAC, WOMAC subscores, and VAS assessment of handicap) between the 100 mg/day diacerein and placebo groups. A trend toward improvement was shown in the other 2 diacerein groups, but this trend was not significant for the primary and secondary criteria.

In the PP population analysis, a significant difference was observed for VAS assessment of pain on movement between each of the 3 diacerein dosages and the placebo group, and improvement of pain was greater when the diacerein dosage was increased (Table 4). No statistically significant differences were found among the 3 diacerein groups. As was the case in the ITT analysis, significant results in favor of the 100 mg/day diacerein group versus the placebo group were shown for the secondary criteria (normalized WOMAC, WOMAC subscores [stiffness and physical function], and VAS assessment of handicap). The same significant trend was observed in the 150 mg/day diacerein group for the normalized WOMAC and the WOMAC stiffness subscore. Similar results were observed when using the AVE analyses for the ITT and PP populations.

Physicians' and patients' global assessment showed greater improvement in the 3 diacerein groups versus the placebo group. However, a significant level

Table 4. Differences in clinical assessment criteria from baseline to week 24 in each of the groups in the per-protocol population*

	Placebo (n = 80)	50 mg/day diacerein (n = 83)	100 mg/day diacerein (n = 80)	150 mg/day diacerein (n = 75)
Pain, VAS, mm	−14.2 ± 19.2	−20.4 ± 18.8†	−23.2 ± 18.2†	−24.7 ± 18.8†
WOMAC, mm	−29.8 ± 46.4	−40.0 ± 48.5	−50.5 ± 46.7†	−46.6 ± 42.5†
Pain	−56.5 ± 83.5	−73.3 ± 86.6	−83.5 ± 83.4	−81.1 ± 80.5
Stiffness	−17.1 ± 40.4	−25.5 ± 42.0	−35.9 ± 39.1†	−33.2 ± 33.9†
Physical	−168.5 ± 257.9	−211.5 ± 270.2	−276.5 ± 274.1†	−239.1 ± 262.8
Handicap, VAS, mm	−14.1 ± 18.2	−15.8 ± 22.6	−22.6 ± 19.9†	−18.8 ± 19.9
Global efficacy assessment, mm				
Investigator	52.3 ± 30.1	62.3 ± 28.3†	58.9 ± 25.0	60.5 ± 30.3
Patient	52.9 ± 30.9	62.7 ± 28.1†	61.1 ± 24.6	61.0 ± 29.3

* Values are the mean ± SD. See Table 1 for definitions.
† $P < 0.05$ versus placebo group.

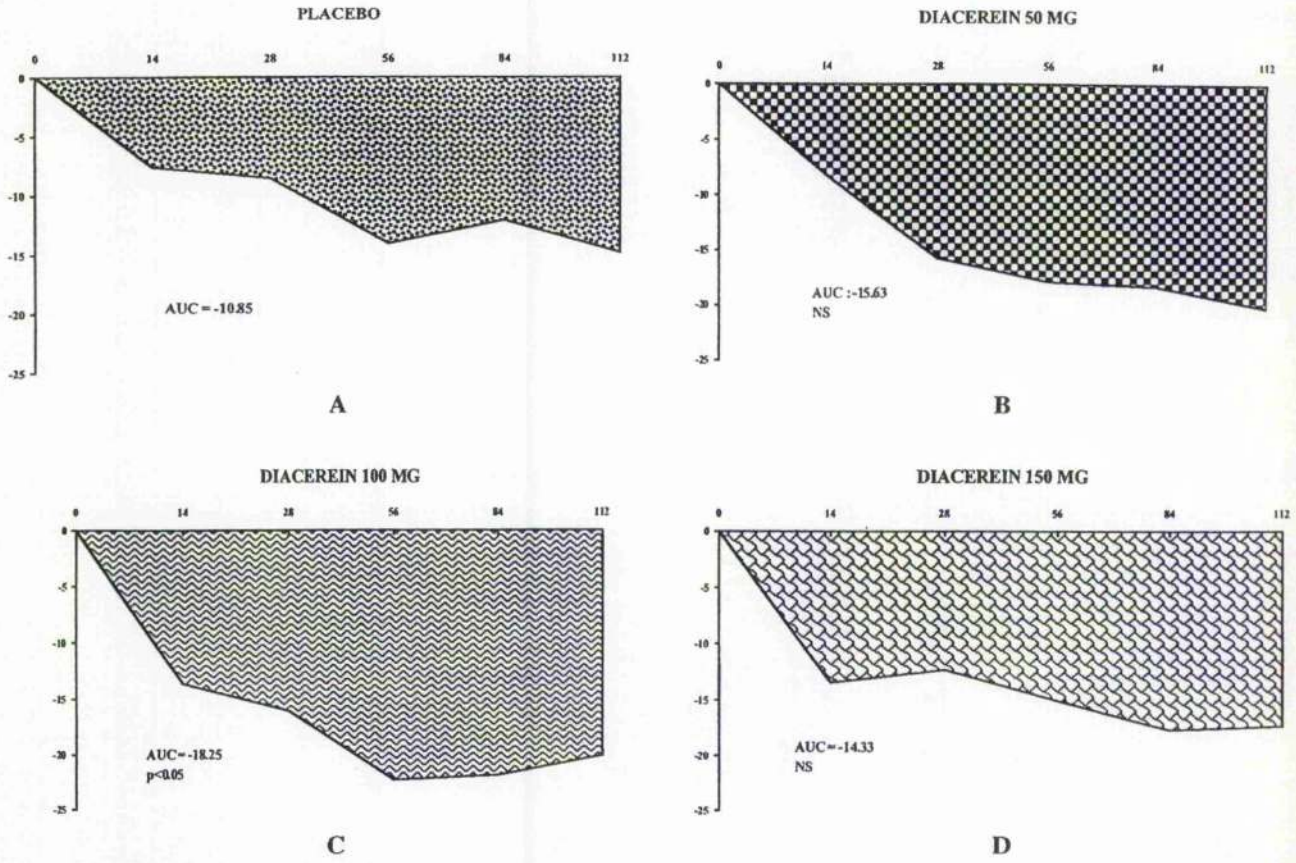


Figure 2. Mean change over time of pain on movement (assessed by visual analog scale) by treatment group for the intent-to-treat (ITT) population. The analysis was performed for the 480 patients who entered the ITT population. Panels A–D represent the area under the curve (AUC) of the pain over time, and normalized values of AUC are presented under the curve. A significant difference ($P < 0.05$) was found for the AUC of pain between the 100 mg/day diacerein group (C) and the placebo group (A). Favorable trends versus placebo were observed for the 50 mg/day diacerein group (B) and the 150 mg/day diacerein group (D), but these trends were not significant (NS).

was reached only in the 50 mg/day diacerein group in the PP population (Table 4).

According to the analysis plan, and in view of the significant results observed for the primary efficacy parameter for the ITT and PP populations, a quadratic model was fitted to estimate the most effective dosage of diacerein. The results indicated a significant fit for the ITT ($P = 0.0255$) (Figure 3) and PP ($P = 0.0007$) (Figure 4) analyses. The best diacerein dosages were estimated to be 90.1 mg/day (ITT population) and 145.3 mg/day (PP population).

An exploratory analysis, based on VAS assessment of pain on movement and performed to determine the onset of action of the 100 mg/day diacerein treatment, indicated that differences from placebo were statistically significant by week 4 in the ITT and PP populations ($P = 0.0338$ and $P = 0.017$, respectively).

Safety. A total of 327 of the 484 randomly assigned patients (67.6%) presented with AEs (Table 5). The proportions of patients who experienced AEs were comparable among the placebo (59.2%), 50 mg/day diacerein (65.1%), and 100 mg/day diacerein (64.0%) groups, while a higher proportion (82.0%) was observed in the 150 mg/day diacerein group. A significantly higher frequency of AEs was observed for the 150 mg/day diacerein group versus the placebo ($P < 0.001$), 50 mg/day diacerein ($P = 0.004$), and 100 mg/day diacerein ($P = 0.003$) groups.

A total of 64 patients (13.2%) discontinued the study due to AEs: 14 of 125 (11.2%), 16 of 126 (12.7%),

11 of 111 (9.9%), and 23 of 122 (18.9%) in the placebo, 50 mg/day diacerein, 100 mg/day diacerein, and 150 mg/day diacerein groups, respectively (Table 2). The main AE was generally mild-to-moderate diarrhea, which occurred in 28.3% of the patients (13.6%, 17.5%, 29.7%, and 53.3% in the placebo, 50 mg/day diacerein, 100 mg/day diacerein, and 150 mg/day diacerein groups, respectively). Diarrhea was considered "severe" in 13 patients in the 150 mg/day diacerein group and in only 2, 1, and 2 patients in the placebo, 50 mg/day diacerein, and 100 mg/day diacerein groups, respectively. Withdrawals due to diarrhea were reported for 12 patients in the 150 mg/day diacerein group compared with 3 patients in each of the other 3 groups. No serious or severe AEs regarding the upper GI tract occurred during the study.

No clinically relevant differences were observed between any of the diacerein groups and the placebo group with regard to vital signs and laboratory analysis (blood and urine). A larger number of patients assessed safety across visits as "good" and "very good" in the 50 and 100 mg/day diacerein groups (86–97% and 77–87%, respectively) than did so in the 150 mg/day diacerein group (63.5–84%). Similar results were observed for the investigators' assessment.

DISCUSSION

The results of this dose-finding study confirm previous findings (12–17) that diacerein is an effective treatment for the signs and symptoms of knee OA, and that based on the results from ITT analysis, the optimal daily dosage is 100 mg/day (50 mg twice daily).

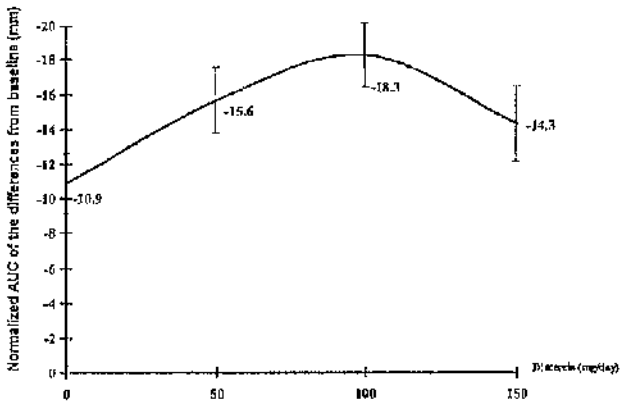


Figure 3. Shape and location of the dose-response curve for visual analog scale (VAS) assessment of pain on movement in the intent-to-treat population (area under the curve [AUC]). The values indicated on the curve correspond to the mean ± SEM values of the AUC of pain on movement (VAS assessment in mm) for each of the tested treatment groups. A significant difference was found between the 100 mg/day diacerein group and the placebo group ($P < 0.05$). A significant fit for the quadratic polynomial model was found ($P = 0.0255$), and the best diacerein dosage was estimated to be 90.1 mg/day.

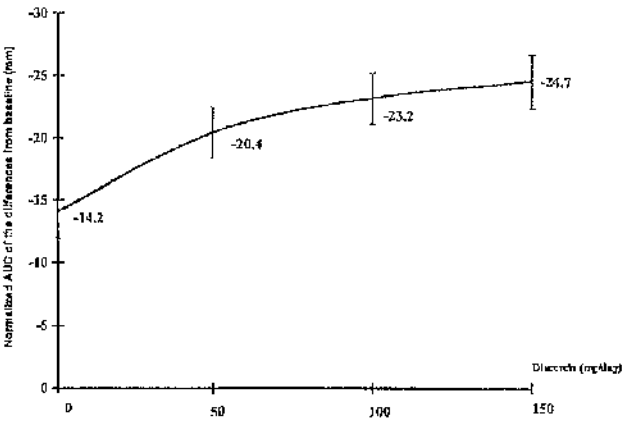


Figure 4. Shape and location of the dose-response curve for visual analog scale (VAS) assessment of pain on movement in the per-protocol population (area under the curve [AUC]). The values indicated on the curve correspond to the mean ± SEM values of the AUC of pain on movement (VAS assessment in mm) for each of the tested treatment groups. A significant difference was found between the placebo group and each of the 3 diacerein groups ($P < 0.05$). A significant fit for the quadratic polynomial model was found ($P = 0.0007$), and the best diacerein dosage was estimated to be 145.3 mg/day.

Diacerein is a drug for the symptomatic treatment of OA that has no inhibitory effects on prostaglandin synthesis. This may explain why diacerein is safe for the upper GI system, as evidenced by the absence of upper GI complaints and findings in clinical trials.

The results of this phase II study indicate that with regard to efficacy, the treatment with diacerein, judged by the primary criterion of VAS assessment of pain on movement, was superior to placebo in all 3 treatment groups, irrespective of the population analyzed (ITT or PP). In the ITT population, a significant difference was found for VAS assessment of pain only for patients receiving 100 mg/day diacerein, while improvement detected in the 2 other diacerein groups was not found to be significant. Moreover, the between-group difference in the improvement of pain level assessed with the WOMAC pain subscore (walking on a flat surface, going up or down stairs, at night while in bed, sitting or lying, and standing upright) was also found to be statistically significant in the 100 mg/day diacerein group for the ITT population. In the PP population, the WOMAC pain subscore showed a favorable trend and approached, but did not quite reach,

Table 5. Adverse events*

	Placebo (n = 125)	50 mg/day diacerein (n = 126)	100 mg/day diacerein (n = 111)	150 mg/day diacerein (n = 122)	Total (n = 484)
Diarrhea	13.6	17.5	29.7†	53.3‡	28.3
Abdominal pain	12.0	16.7	18.0	25.4†	18.0
Soft stools	4.8	8.7	9.0	11.5	8.5
Headache	8.8	5.6	8.1	9.8	8.1
Nausea	4.0	7.1	9.0	11.5§	7.9
Musculoskeletal pain	8.8¶	6.4	0.9	3.3	5.0
Weariness	2.4	5.6	1.8	6.6	4.1
Dyspepsia	3.2	2.4	6.3	4.9	4.1

* Values are the percentage of patients experiencing the event. Only events that occurred in ≥5% of any group are shown.
† $P < 0.01$ versus placebo.
‡ $P < 0.001$ versus placebo.
§ $P < 0.05$ versus placebo.
¶ $P < 0.01$ versus 100 mg/day diacerein.

significance for the 100 mg/day diacerein group. This can be explained by the number of patients who withdrew from the study prematurely, dramatically decreasing the power of the study.

The large number of withdrawals in the 150 mg/day diacerein group (mainly at the beginning of the treatment period) was due to a high incidence of AEs that was probably related to the treatment, and may explain the absence of a dose-effect relationship in the ITT population. Because of the delay of action of diacerein's effect, the LOCF procedure used in the present analysis disadvantaged the 150 mg/day diacerein group by carrying forward high values for VAS assessments of pain, close to those at baseline. This hypothesis is supported by the results obtained in the PP population, where statistically significant differences were shown for the 3 groups receiving diacerein. Furthermore, the shape and location of the dose-response curve in the PP population indicated that the VAS assessment of pain significantly decreased in a dose-dependent manner. The efficacy of treatment with diacerein was also supported by the analysis of the secondary endpoints (WOMAC, WOMAC subscores, and VAS assessment of handicap), which showed statistically significant differences compared with placebo for the group receiving 100 mg/day diacerein in the ITT population. According to the shape and location of the dose-response curves, the best daily dosage of diacerein was very close to 100 mg/day (ITT population) (calculation from the quadratic fit yielded 90.1 mg/day; $P = 0.0255$).

The analysis of the onset of the beneficial effect of diacerein showed a delay of action, with a statistically significant effect starting from week 4, particularly in the group receiving 100 mg/day diacerein, in both the ITT ($P = 0.034$) and PP ($P = 0.017$) populations. These results are in accord with the findings of previous clinical studies (12,14,15). Moreover, unlike the case with NSAIDs, sustained pain relief lasting 1–2 months was observed after discontinuation of diacerein, showing a carryover effect of this treatment (13,15,17). This could obviously represent a significant advantage of diacerein over NSAID treatment in patients who must stop taking their medication for several consecutive days.

Regarding the safety data globally, the number of patients reporting AEs was found to differ significantly among the 4 groups ($P < 0.001$). This difference was due to the 150 mg/day diacerein group, in which the number of AEs was significantly higher ($P < 0.01$) compared with the placebo, 50 mg/day diacerein, and 100 mg/day diacerein groups (1.7-fold, 1.5-fold, and 1.6-fold, respectively). Furthermore, the AEs occurring in the group receiving 150 mg/day diacerein were more likely to be judged to be "related" to the study treatment. As expected from the results of previous clinical trials, changes in bowel habits (diarrhea, soft stools, and abdominal pain) were the most frequently reported symptoms in the diacerein-treated groups, with a clear dose-effect relationship. These side effects were generally classified as "mild" by the patients.

The first possible explanation for this kind of side effect with diacerein is that it is a drug class effect (29). Another possible, although quite hypothetical, explanation is that since diacerein has been shown to be capable of inducing prostaglandin synthesis, it may be that a local increase in prostaglandins can lead to an increase in gut motility and thus to diarrhea (30). No severe or serious

AEs concerning the upper GI tract, such as gastric or duodenal ulcers, were reported during the trial. These findings are reassuring and are consistent with those of postmarketing surveillance (since 1994) in several countries (including France) where diacerein is available, as well as with the results of previous clinical trials.

Diacerein is therefore a safe alternative to NSAIDs for the treatment of OA. In a 2-month, double-blind, 2 × 2 factorial plan study including 288 patients with hip OA, 100 mg/day diacerein was compared with a placebo, an NSAID (20 mg/day tenoxicam), and a combination of the same dosages of diacerein and tenoxicam (12). The improvement of pain on movement and algofunctional Lequesne index showed a significant difference versus placebo for the tenoxicam and combination groups after 2 weeks of treatment. Similar improvement was observed for the diacerein group, becoming significant after 6 weeks of treatment. No differences were observed among the tenoxicam, combination treatment, and diacerein groups. In a second double-blind study (17), 95 patients with knee or hip OA were given either 100 mg/day diacerein or 750 mg/day naproxen for 2 months. Similar significant pain improvement was found in both treatment groups.

The results observed in recent clinical trials in the treatment of OA involving NSAIDs showed that the magnitude of pain improvement in patients (versus placebo, assessed with a VAS) was comparable with that observed with 100 mg/day diacerein (−7.4 mm to −10.0 mm), as follows: 15 mg/day meloxicam, −6.1 mm (31); 200 mg/day nimesulide, −6.0 mm (32); and 200 mg/day celecoxib, −7.6 mm to −12.2 mm (33). One should, however, exercise caution in making this comparison, since the assessments of pain were not necessarily equivalent among studies (different pain condition assessments). These data nevertheless provide some valuable comparative information with regard to diacerein's effect.

In terms of safety, NSAIDs are known to have deleterious effects on the gastric mucosa, including complications such as peptic ulcers, gastritis, perforation, and gastric hemorrhage; these effects are related to their inhibitory action on COX, with a consequent reduction in prostaglandin synthesis. The new generation of NSAIDs, such as the COX-2 inhibitors, seems to have fewer upper GI tract side effects (34,35). However, further clinical trials are being conducted to fully assess the safety profile of these drugs. Diacerein, an inhibitor of IL-1 production, has a mode of action which differs in many ways from that of NSAIDs. Given that this drug has no inhibitory effects on COX, it offers a safe and effective treatment for the symptoms of OA.

Based on its clinical efficacy for the signs and symptoms of OA, its structure modification potential, and its safety profile, diacerein constitutes a novel approach to the treatment of OA. It can be envisioned as a treatment for the short- and long-term management of this disease.

ACKNOWLEDGMENTS

The authors would like to thank the following investigators for their work, which made this study possible: From Israel, D. Caspi, O. Elkayam, D. Levralovski, D. Paran, T. Reitblat, M. Tishler (Tel Aviv); A. Burstein, A. Gurman-Balbir, R. Mader, M. Rozenbaum, D. Schupira (Haifa); A. Rubinow (Jerusalem); C. Brickman, O. Schoenberg, A. Tanay (Holon); M. Abu-Shakra, D. Buskila (Beer-Sheva); L. Green, Y. Fridman (Rehovot); D. Guedj, Y. Molad (Petah-Tiqva); P. Langevitz (Tel-Hashomer); C. Vigderer, A. Yaretzky (Kfar Saba); from Canada, T. P. Anastassiades (Kingston); N. Attié (Saint-Leonard); M. J. Bell (Toronto); N. Bellamy (London); W. G. Bensen (Hamilton); M. A. Fitzcharles, J. P. Raynauld, É. Rich, H. Tannenbaum, M. Zummer (Montreal); A. Lussier (Sherbrooke); J. Carter-Thorne (Newmarket).

REFERENCES

1. Spencer CM, Wilde MI. Diacerein. Drugs 1997;53:98–108.
2. Mian M, Trombi I, Rosini S, Benetti D, Caracciolo F, Carulli G, et al. Experimental studies on diacerhein: effects on the phagocytosis of neutrophil cells from subcutaneous carrageenan induced exudate. Drugs Exp Clin Res 1987;13:695–8.
3. Pelletier JP, Mineau F, Fernandes JC, Duval N, Martel-Pelletier J. Diacerhein and rhein reduce the interleukin-1 beta stimulated inducible nitric oxide synthesis level and activity while stimulating cyclooxygenase-2 synthesis in human osteoarthritic chondrocytes. J Rheumatol 1998;25:2417–24.
4. Martel-Pelletier J, Mineau F, Jolicoeur FC, Cloutier JM, Pelletier JP. In vitro effects of diacerhein and rhein on interleukin-1 and tumor necrosis factor-alpha systems in human osteoarthritic synovium and chondrocytes. J Rheumatol 1998;25:753–62.
5. Yaron M, Shirazi I, Yaron I. Anti-interleukin-1 effects of diacerein and rhein in human osteoarthritic synovial tissue and cartilage cultures. Osteoarthritis Cartilage 1999;7:272–80.
6. Boittin M, Redini F, Loyau G, Pujol JP. Effet de la diacerhéine (ART 50) sur la synthèse de la matrice et la sécrétion de collagénase par des chondrocytes articulaires de lapin en culture. Rev Rhum 1993;60:S68–76.
7. Mian M, Benetti D, Rosini S, Pantozzi R. Rhein reduces proteoglycan loss during the autolytic breakdown of cultured cartilage. Int J Tissue React 1989;11:117–22.
8. Smith GN Jr, Myers SL, Brandt KD, Mickler EA, Albrecht ME. Diacerhein treatment reduces the severity of osteoarthritis in the canine cruciate-deficiency model of osteoarthritis. Arthritis Rheum 1999;42:545–54.
9. Ghosh P, Xu A, Hwa SY, Burkhardt D, Little C. Evaluation des effets de la diacerhéine dans un modèle ovin d'arthrose. Rev Prat 1998;48 Suppl 17:S24–30.

10. Bendele AM, Bendele RA, Hulman JF, Swaun BP. Effets bénéfiques d'un traitement par la diacerhéine chez des cobayes atteints d'arthrose. Rev Prat 1996;46 Suppl 6:S35–9.
11. Mazieres B. Diacerhein in a post-contusive model of osteoarthritis: structural results with "prophylactic" and "curative" regimens [abstract]. Osteoarthritis Cartilage 1997;5 Suppl A:73.
12. Nguyen M, Dougados M, Berdah L, Amor B. Diacerhein in the treatment of osteoarthritis of the hip. Arthritis Rheum 1994;37: 529–36.
13. Lequesne M, Berdah L, Gerentes I. Efficacité et tolérance de la diacerhéine dans le traitement de la gonarthrose et de la coxarthrose. Rev Prat 1998;48 Suppl 17:S31–5.
14. Ascherl R. Longterm treatment of osteoarthritis of the knee: a six months placebo-controlled clinical trial with diacerein [abstract]. Osteoarthritis Cartilage 1994;2 Suppl 1:18.
15. Vidal JP. Etude pragmatique de la diacerhéine: résultats sur les signes cliniques, la qualité de vie et les coûts médico-économiques. Rev Prat 1997;47 Suppl 17:S39–45.
16. Fagnani F, Bouvenot G, Valat JP, Bardin T, Berdah L, Lafuma A, et al. Medico-economic analysis of diacerein with or without standard therapy in the treatment of osteoarthritis. Pharmacoeconomics 1998;13:135–46.
17. Marcolongo R, Fioravanti A, Adami S, Tozzi E, Mian M, Zampieri A. Efficacy and tolerability of diacerhein in the treatment of osteoarthrosis. Curr Ther Res 1988;43:878–87.
18. Petrillo M, Montrone F, Ardizzone S, Caruso I, Bianchi-Porro G. Endoscopic evaluation of diacetylrhein induced gastric mucosal lesions. Curr Ther Res 1991;49:10–5.
19. Bellary SV, Isaacs PE, Lee FI. Upper gastrointestinal lesions in elderly patients presenting for endoscopy: relevance of NSAID usage. Am J Gastroenterol 1991;86:961–4.
20. Kellgren JH, Lawrence JS. Radiological assessment of osteoarthrosis. Ann Rheum Dis 1957;16:494–502.
21. Altman R, Asch E, Bloch D, Bole G, Borenstein D, Brandt K, et al. Development of criteria for the classification and reporting of osteoarthritis: classification of osteoarthritis of the knee. Arthritis Rheum 1986;29:1039–49.
22. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988;15:1833–40.
23. Choquette D, Bellamy N, Raynauld JP. A French-Canadian version of the WOMAC osteoarthritis index [abstract]. Arthritis Rheum 1994;37 Suppl 9:S226.
24. Wigler I, Neumann L, Yaron M. Validation study of a Hebrew version of WOMAC in patients with osteoarthritis of the knee. Clin Rheumatol 1999;18:402–5.
25. Lequesne M, Samson M, Bakouche S. Utilisation d'une échelle visuelle analogique (EVA) pour l'autoappréciation du handicap dans les coxopathies de l'adulte [abstract]. Rev Rhum 1991;58:694.
26. Duruoz MT, Poiraudeau S, Fermanian J, Menkes CJ, Amor B, Dougados M, et al. Development and validation of a rheumatoid hand functional disability scale that assesses functional handicap. J Rheumatol 1996;23:1167–72.
27. Dawson JD. Comparing treatment groups on the basis of slopes, areas-under-the-curve, and other summary measures. Drug Information J 1994;28:723–32.
28. Fleis JL. Special cases of the parallel groups study. In: Fleiss JL, editor. The design and analysis of clinical experiments. New York: John Wiley & Sons; 1985. p. 91–119.
29. Leng-Peschlow E. Effect of sennosides and related compounds on intestinal transit in the rat. Pharmacology 1988;36 Suppl 1:40–8.
30. Yagi T, Miyawaki Y, Nishikawa T, Yamauchi K, Kuwano S. Involvement of prostaglandin E-like material in the purgative action of rhein anthrone, the intraluminal active metabolite of sennosides A and B in mice. J Pharm Pharmacol 1988;40:27–30.
31. Lund B, Distel M, Bluhmki E. A double-blind, randomized, placebo-controlled study of efficacy and tolerance of meloxicam treatment in patients with osteoarthritis of the knee. Scand J Rheumatol 1998;27:32–7.
32. Bourgeois P, Dreiser RL, Lequesne MG, Macciocchi A, Mouti T. Multi-centre double-blind study to define the most favourable dose of nimesulide in terms of efficacy/safety ratio in the treatment of osteoarthritis. Eur J Rheumatol Inflamm 1994;14:39–50.
33. US Food and Drug Administration Center for Drug Evaluation and Research. Celecoxib: medical review application number NDA 20-998. Accessed 10/22/99. URL: http://www.fda.gov/cder/foi/nda/98/20998.htm.
34. Langman MJ, Jensen DM, Watson DJ, Harper SE, Zhao PL, Quan H, et al. Adverse upper gastrointestinal effects of rofecoxib compared with NSAIDs. JAMA 1999;282:1929–33.
35. Simon LS, Weaver AL, Graham DY, Kivitz AJ, Lipsky PE, Hubbard RC, et al. Anti-inflammatory and upper gastrointestinal effects of celecoxib in rheumatoid arthritis: a randomized controlled trial. JAMA 1999;282:1921–8.

The Journal of Continuing Education in the Health Professions, Volume 20, pp. 52–61. Printed in the U.S.A.

Original Article

Continuing Medical Education-Driven Skills Acquisition and Impact on Improved Patient Outcomes in Family Practice Setting

Nicholas Bellamy, MD, MS, FRCP(Glas), FRCP(Edin), FRCPC, FACP, FAFRM, Laurence D. Goldstein, BSc, and Rory A. Tekanoff, PEng(Dipl), APMR

Abstract

Background: An abundance of educational theory, design, and delivery of continuing medical education (CME) learning interventions, including their impact on learners, are described in the health and social sciences literature. However, establishing a direct correlation between the acquisition of new skills by learners and patient outcomes as a result of a planned CME learning intervention has been difficult to demonstrate.

Methods: The learning intervention described here tested the impact of an injection skills-acquisition program for family physicians treating osteoarthritis of the knee by measuring patient outcomes using the pain and function subscales of the Western Ontario and McMaster (WOMAC) 3.0 osteoarthritis index, a standardized and fully validated patient-centered outcome measurement. It was hypothesized that patients of family physicians who participated in this skills-acquisition CME program would benefit from treatment administered by their physician during the time between injection skills acquisition to 6 weeks post-injection. Inclusion of a validated health status measure administered pre- and post-injection in addition to more traditional faculty and participant program evaluations was deemed necessary to test this hypothesis. Rheumatology, orthopedic surgery, and family medicine specialists from across Canada were invited to contribute to the planning, curriculum elaboration, and delivery of the viscosupplement injector preceptorship (VIP) program. Thirty-nine orthopedic and rheumatology specialists agreed to serve as expert faculty and participated in training 474 Canadian family and general practitioners over 8 months. The learning intervention involved a review of pertinent literature by a local preceptor and a summary of recommendations of the planning committee, followed by demonstration of injector skills and then supervised practice with patients, who received hylan G-F 20 (Synvisc™, Ridgefield, NJ) usually in the offices of the family physicians. The pain and function subscales of the WOMAC 3.0 questionnaire were self-administered to each patient in their physician's office, prior to receiving their joint injection and again at or near 6-weeks post-injection. Data were analyzed in the Department of Epidemiology and Biostatistics at The University of Western Ontario, London, ON.

Results: Clinically important statistically significant improvements in pain and physical function were noted in patients who received viscosupplementation treatment from family physicians who had recently acquired the necessary injection skills. Approximately three-quarters of the

Dr. Bellamy: Professor and Director of the Centre of National Research on Disability and Rehabilitation Medicine (CONROD), University of Queensland, Faculty of Health Sciences, Royal Brisbane Hospital, Brisbane, Australia; *Mr. Goldstein:* President, KARMA® Clinical Relations Canada Inc., Toronto, ON; and *Mr. Tekanoff:* Senior Manager, Continuing Health Education, Parke-Davis Canada, Scarborough, ON.

Reprint requests: Nicholas Bellamy, MD, MSc, FRCP(Glas), FRCP(Edin), FRCPC, FACP, FAFRM, University of Queensland, Faculty of Health Sciences, Royal Brisbane Hospital, Queensland 4029, Australia.

patients experienced a reduction in pain and an improvement in physical function of at least 20%.

Implications: These results suggest a positive relationship between acquisition of a new skill by learners and improved patient outcomes as a result of this planned CME learning intervention.

Key Words: Continuing medical education (CME), family physicians, injection, measurement devices, patient outcomes, skills acquisition, validation, viscosupplementation, WOMBAT, WOMAC

An environmental scan of the continuing medical education (CME), rheumatology, and orthopedic literature and input from expert preceptors from across Canada revealed there is no standardized country-wide undergraduate or postgraduate curriculum or other well-defined educational process for acquiring joint injection skills.[1] For most family physicians, these skills seem to be acquired on a rheumatology or orthopaedic rotation as an undergraduate, intern, or resident, or from a colleague while in practice. A needs assessment conducted among 44 family physicians in southwestern Ontario indicated that 59% were interested in acquiring knowledge about joint injection, of which half expressed particular interest in viscosupplementation. This article reports the development of a knowledge- and skills-based CME program. The impact of the program on participating family practitioners and on the patients they subsequently treated with their newly acquired skill was also evaluated.

Methods

A curriculum-elaboration meeting was organized, involving seven medical specialists in orthopedic surgery, rheumatology, and family medicine; these specialists also formed a steering committee. This group was to provide professional input in the development of a curriculum for injection training of family physicians and elaborate the scope and dimension of a suitable learning intervention, including an instruction manual and participant workbook along with other relevant CME materials. A 1-day meeting was held to define target audience needs, training group size, faculty qualifications, training locations for physician instruction, patient qualifiers, evaluative mechanisms, and incentives. It was also agreed that viscosupplementation with hylan GF 20 (Synvisc™, Ridgefield, NJ), an injectable form of osteoarthritis therapy, would be used in the training of family physicians.

A viscosupplement injector preceptorship (VIP) program skills acquisition manual (SAM) was developed with input and review from the steering committee. The SAM was to serve as both a guide for expert faculty and a resource for participants: it included information on basic anatomy of the knee, diagnosis of knee osteoarthritis, treatment guidelines for knee osteoarthritis, and patient selection criteria for viscosupplementation with Synvisc™. Sections addressing mode of action, clinical and adverse effects, warnings and contra-indications to Synvisc™ treatment, and practical pointers were also included. Issues concerning use of injectable corticosteroids were also addressed for the same areas as for viscosupplementation. The remainder of the manual provided practical tips for giving injections and for the prevention and management of adverse reactions, and troubleshooting guidelines. The manual also included samples of all evaluative mechanisms (pre- and post-injection pain and function subscales of the Western Ontario and McMaster [WOMAC] 3.0 questionnaire, faculty expert preceptor-program evaluation, and participant-

preceptor program evaluation questionnaires) and an extensive reference list of peer-reviewed literature. The WOMAC index is widely used, and is a valid, reliable, and responsive self-administered tridimensional health status measure for knee and hip osteoarthritis studies, available in visual analogue and adjectival formats in over 30 different languages.[2]

Faculty trainers (expert preceptors) were contacted by a third party (KARMA® Clinical Relations Canada Inc.) to participate in the program. Faculty identification was based on market research and intelligence provided by the supporting pharmaceutical and device manufacturers, and used the following criteria:

- Specialist or family physicians who have particular skills in injection techniques, primarily in the knee;
- Physicians who possess advanced knowledge of principles of viscosupplementation and its mode of action and success rates, and have used viscosupplementation in the past month;
- Physicians who have successful experience with Synvisc™;
- Physicians who are local experts in osteoarthritis of the knee as evaluated by local family practitioners; and
- Physicians who are interested in CME and in teaching other physicians and health care professionals.

Forty potential expert preceptors were contacted by telephone and interviewed to determine their suitability, interest, and availability to participate in the learning intervention; 39 agreed to participate. Each subsequently met with members of the steering committee either personally or by conference call to discuss the program. Each received a SAM prior to follow-up contact to familiarize themselves with the instruction materials that participants would receive before their training session.

Participants (family practitioners) were identified and contacted by representatives of the pharmaceutical and device manufacturers, and were invited to attend the VIP sessions by the following criteria, evaluated in a personal interview:

- The family practitioner expressed an avid interest in acquiring joint-injection techniques to the representative;
- Viscosupplemenation was not currently used for osteoarthritis therapy in the family practitioner's practice, primarily due to lack of expertise in joint injection;
- The family practitioner was willing to devote 4 hours to participating in a training session;
- The practitioner could provide a patient with osteoarthritis of the knee(s) who was amenable to joint injection;
- The practitioner would participate in the evaluation process (pre- and post-injection WOMAC 3.0 questionnaires); and
- The family practitioner would continue self-directed use of the injection skill to maintain competency.

Representatives then organized a training session for three to five family practitioners and a local expert preceptor; this is an effective format to enhance learning.[3] Each session was conducted in the clinical practice of the participants, which, although variable, always was one of the following: the preceptor's offices or group-practice clinic; the expert preceptor's private practice, or a hospital. A typical training session consisted of a small-group interactive learning session followed by live patient injection-technique demonstration and practice, all taking place over approximately 4 hours. Each family physician was required to bring one patient with osteoarthritis of the knee(s) and the patient's x-rays to the training session. Prior to injection, participants supervised the administration of pre-injection WOMAC 3.0 questionnaires to their patients and the expert preceptor re-examined each patient and confirmed the

osteoarthritis diagnosis and suitability for viscosupplementation therapy. The practical work began when the expert preceptor determined that the group felt ready to begin injecting. Each family practitioner injected their own patient under the supervision of the expert preceptor, while being scrutinized by their peers. Since a full course of Synvisc™ requires three intra-articular injections administered 1 week apart, the preceptor was available to the family practitioner should difficulty be encountered during subsequent injections. In only 9 of 445 (2%) cases did a family practitioner contact the expert preceptor for additional training or to request that the expert preceptor perform the follow-up injections on their patient. An average of four family practitioners participated in each training session, with 96 individual sessions completed in 6 months across Canada.

A second objective of this study was to validate a patient global assessment question for future incorporation into a modified WOMAC 3.0 index, that we have provisionally termed the Western Ontario Measurement Battery (WOMBAT 3.0). If successfully validated, the WOMBAT 3.0 would contain the WOMAC pain and physical function subscales, and a patient global assessment of knee osteoarthritis subscale. In contrast to the WOMAC 3.0, the WOMBAT 3.0 would not contain a stiffness subscale. This modification was to accommodate recommendations made at the OMERACT III Conference[4] and in the Ostearthritis Research Society guidelines document[5] in which pain, function, and patient global assessment (but not stiffness) were established by international consensus as core set clinical variables for future osteoarthritis studies. The WOMAC 3.0 and the patient global assessment question were prepared in adjectival (Likert) format, in both English and French for Canada, and combined in a single questionnaire, hereafter referred to as the WOMAC/PGA questionnaire.

Data were coded, entered, and analyzed in the Department of Epidemiology and Biostatistics at the University of Western Ontario using the SAS program.[6] Descriptive statistics were used to characterize responses to the three questionnaires. The statistical significance of the treatment effect was evaluated using both parametric (Students *t* test) and nonparametric (Wilcoxon Signed Rank test) methods. Previous comparisons of these two approaches using the WOMAC Index has not shown important differences in levels of significance or data interpretation.

In order to validate the patient global assessment question, the approach captured by the OMERACT Filter was used.[7] The OMERACT Filter for selecting outcome measures places emphasis on those that fulfill criteria for truth (validity), sensitivity (responsiveness + reliability), and feasibility. Validity was assessed by testing convergent construct validity between patient global assessment scores and WOMAC pain and function subscale scores. Sensitivity was evaluated by comparing post-injection and pre-injection patient global assessment scores: feasibility was evaluated by observing if completed WOMAC questionnaires were accompanied by completed patient global assessment questions.

Results

Only 445 five patients received injections, since 29 of the 474 physicians were unable to supply a patient but had not stated this in the interview selection process (completion rate = 94%). Of the 890 potentially available WOMAC/PGA questionnaires, 602 were returned sometime after the first injection: of these, there were 163 complete pairs (i.e., preaccompanied by post) that were used in the analyses reported. Of the remainder, 115 had only a pre- and 25 only a post-injection assessment, 18 contained data from different knees at the two assessment points, and in 118 (59 pairs), the post-injection assessment was made less than 21 or more than 84 days after the pre-injection questionnaire was completed. Although all post-injection assessments should have been completed at 6 weeks, there was considerable variation in when post-injection assessments were made (Figure 1). We elected to restrict the data analysis to

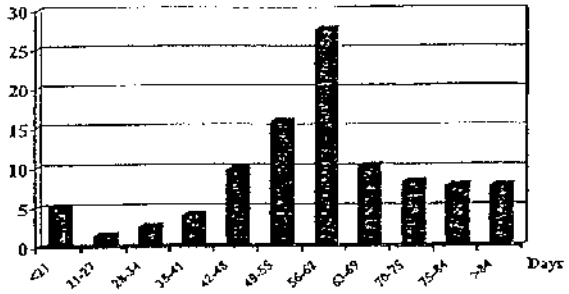


Figure 1 Time to completion of post-injection WOMAC 3.0.

subjects who had completed their post-injection assessments between 3 and 12 weeks. Three weeks is the first week after completion of the series of three injections, and 12 weeks is the point used in several published studies to evaluate the early response to Synvisc™.

Data from 163 subjects were used in the analysis. Pre- and post-injection WOMAC scores and associated change scores for the two subscales are presented in Table 1. These improvements were clinically important and statistically significant for both WOMAC 3.0 subscales (pain = physical function at $p < .001$, by both parametric and nonparametric analyses). Notwithstanding the current lack of responder criteria for osteoarthritis knee studies, patients were classed as responders if they fit either of the following two definitions: 20% or more reduction in pain and 20% or more reduction in pain as well as 20% or more improvement in physical function. Seventy-four percent of participants were responders by the first definition and 73% by the second.

More than 94% of family practitioners agreed or strongly agreed that the VIP program was practical and relevant, met their objectives and expectations, was credible and well organized, that time and interaction were adequate, that the preceptor was knowledgeable, and that they now felt comfortable with the procedure and would consider viscosupplementation as a treatment option for osteoarthritis knee patients in their practice (Table 2). Almost 100% of preceptors agreed or strongly agreed that the program was practical and relevant, met their objectives and expectations, was credible and well organized, that time and interaction were adequate, and that they would participate in future VIP programs.

With respect to the validation of the patient global assessment question, there was a strong positive correlation between patient global assessment scores and WOMAC pain and function scores. For pre-injection, pain had an $r = 0.59$ and function had an $r = 0.62$ at $p < .001$; post-injection pain had an $r = 0.79$ and function had an $r = 0.77$ at $p < .001$; the change score for pain had an $r = 0.69$ and function, an $r = 0.71$ at $p < .001$.

Table 1 Clinical Profiles Pre- and Post-Injection with Synvisc™

	Variable	n	Mean	SD	Min.	Max.
Pre-injection	WOMAC pain	163	10.55	3.45	1	20
	WOMAC function	163	37.36	11.67	8	68
	Global assessment	157	2.94	0.79	1	4
Post-injection	WOMAC pain	161	6.04	4.5	0	20
	WOMAC function	163	23.18	14.36	0	66
	Global assessment	161	1.71	1.14	0	4
Pre- to post-difference	WOMAC pain	161	4.50*	4.16	–11	18
	WOMAC function	163	14.18*	14.32	–37.69	54
	Global assessment	156	1.21*	1.19	–3	4

* $p < .001$, by Student's *t*-test and Wilcoxon Signed Rank test.

Table 2 VIP Program Evaluation Summary

Ranking		Neutral (%)	Agree (%)	Strongly Agree (%)
Practice relevancy	EP	–	26.0	74.0
	FP	0.6	29.2	69.6
Met course objectives	EP	–	26.0	74.0
	FP	1.9	26.1	71.4
Met personal expectations	EP	–	32.0	68.0
	FP	2.5	27.5	69.4
Credible	EP	..	42.0	58.0
	FP	3.1	31.7	64.6
Well organized	EP	5.0	37.0	58.0
	FP	1.9	27.3	69.6
Adequate time for learning	EP	–	26.0	74.0
	FP	2.5	25.8	70.4
Adequate interaction with expert and peers	EP	–	21.0	79.0
	FP	0	20.1	78.6
Participate as expert again	EP	–	21.0	79.0
	FP	–	–	–
Learning objectives met	EP	–	–	–
	FP	0.6	25.0	73.7
Knowledgeable and skilled expert preceptor	EP	–	–	–
	FP	0.6	13.8	84.9
Comfortable with repeating procedure	EP	–	–	–
	FP	4.6	28.7	65.7
Will use viscosupplement in practice	EP	..	–	–
	FP	–	38.3	60.7

EP = expert preceptor, FP = family practitioner.

Pre-injection and post-injection patient global assessment scores and the associated change scores relating to responsiveness are illustrated in Table 1.

The improvements noted were clinically important and statistically significant ($p < .001$ by both parametric and nonparametric analyses). With respect to feasibility, all completed WOMAC questionnaires were accompanied by completed patient global assessment questions.

Discussion

The majority of previous CME studies have assessed the consequence of the CME intervention at the level of the motivation, knowledge, or intention to change behavior by the learner.[8–12] While these are useful endpoints from an educational standpoint, they leave unanswered the more important question of whether the CME program had a meaningful and beneficial impact on the health status of patients subsequently treated by those who participated. This tendency to measure more-proximal endpoints is understandable, since the measurement of clinical consequence is both complex and costly. Furthermore, it is difficult to directly attribute alterations in the health status of patients to the learning intervention; this may have deterred some previous investigators from pursuing the more important distal endpoints. Clearly, attendees at CME events are to some

extent self-selected by motivation, need, and ambition, and such individuals are not readily randomized by whether they do or do not attend the CME event.[13,14] Moreover, once attendees return to their practices, it is no longer possible to randomize their patients and examine differential effects of being treated by their own physician with versus without the recently acquired skill.[15–18]

In this study, not only the experience of the preceptors and learners but also the changing health status of those patients whom they treated immediately after acquiring the skills necessary to perform viscosupplementation were evaluated. The preceptors were clearly satisfied with the educational experience even prior to observing a beneficial effect on the patients they subsequently injected. It is noteworthy that, following review of the SAM and completion of a supervised injection of Synvisc™, almost all family practitioners felt comfortable about viscosupplementation and would consider its use in future management of osteoarthritis knee patients. For the family practitioners, this represented a relatively small time commitment to acquire a skill of general value in the management of osteoarthritis and delivery of intra-articular therapy. It also permitted skills acquisition to occur in a clinical environment supportive for both the family practitioner and participating patients. The preceptors were often teaching in locations remote from their practices and on patients they had not previously met or examined. That the preceptors were also satisfied with their involvement in the VIP programs and would participate in future programs underscores the success of the intervention.

This is particularly remarkable given the inherent difficulty of maintaining consistency in delivering a learning intervention at multiple sites in different geographic areas with regional variations in health care systems, and involving faculty and learners with different medical specialty backgrounds. The uniformly high level of satisfaction expressed by faculty and learners can be attributed to the planning, design, and delivery approaches employed in this intervention. Variations of several adult learning principles and other approaches described in the literature were adopted and applied,[19–21] and may be summarized as follows:

1. A multi-stakeholder approach, receiving input and validation by faculty and learners at each stage of the development and delivery processes;

2. Multiple learning devices versus reliance on a single educational event;

3. Delivery in or close to the community in which the learners practice;

4. Learners involving their own patients rather than artificial models;

5. Small learning groups; and

6. Self-assessment and immediate feedback to faculty and learners from their own observations of patient outcomes provided by their administration of the pre- and post-injection WOMAC 3.0 questionnaires.

In assessing patient outcomes, it is important to use measures that are valid, reliable, and responsive. The WOMAC osteoarthritis index is one such measure, and has been extensively used in over 50 countries throughout the world.[22,23] In this study, clinically important improvements were noted in WOMAC pain and function scores. Furthermore, while there is currently no internationally accepted definition of responder criteria for osteoarthritis knee studies, a cutpoint used in rheumatoid arthritis studies[24] was borrowed and three-quarters of patients were observed to experience a clinically meaningful response in both pain and function following viscosupplementation.

Furthermore, the patient global assessment question used in this study was shown to be valid, sensitive to change, and feasible, thus fulfilling the requirements of the OMERACT filter. It is therefore proposed that the patient global assessment

be used to supplement the standard WOMAC 3.0 index to create a WOMAC 4.0, or that it replaces the stiffness subscale in the WOMAC 3.0 to create a modified index termed the WOMBAT 3.0 that meets OMERACT/OARSI guidelines.

Potential limitations of the study merit consideration. In general, bias may occur as control over experimental conditions diminishes. Clinical benefit was observed among patients treated by participants in the VIP program. Since the program was delivered as a package, the relative contribution of its different components cannot be discerned. However, the combination of the SAM and the experienced preceptors provided optimum conditions for small-group learning and for practicing a newly acquired skill. This was an open study in which expectation bias both by the family practitioner and patient could modulate the response, potentially in a favorable direction. For example, the family practitioner could have presented the possible benefits to patients in an enthusiastic way, and patients who elected to participate might be self-selected on that basis. However, double-blind randomized placebo-controlled trials of Synvisc™ have demonstrated the intrinsic efficacy of viscosupplementation,[25] which has been substantiated in controlled trials of nonsteroidal class agents[26] as well as in open studies,[27] indicating that while the response may be modulated by expectation in some patients, it does not account for the improvement in health status observed.

Some procedures required for this study were more commonly used in clinical trials based in academic centers. The requirements for patient selection were detailed in the SAM, while the verification of a diagnosis of knee osteoarthritis was performed by the expert preceptor based on a personal interview and examination of the knee and accompanying radiographs. Of the 445 patients who participated, complete WOMAC/patient global assessment data within the 3- to 12-week period were available on 163, although some data were available on 380 patients. These protocol violations and losses to follow-up are likely attributable to the absence of a clinical research organization monitoring the data acquisition, a contingency strongly recommended for future family practitioner-based studies of this type.

The use of a self-administered outcome measure (WOMAC 3.0) in this study obviated any family practitioner-associated positive reporting. While the global question in the WOMBAT is new, this question is valid, responsive, and feasible, and may be used in future osteoarthritis knee studies either within the WOMBAT 3.0 or as a supplementary question within a WOMAC 4.0 index.

Conclusion

The VIP program was successful in training family practitioners to apply a safe and effective intra-articular therapy in patients with knee osteoarthritis. The most important design element was longitudinal evaluation of those patients who were recipients of a family practitioner's newly acquired skill. The clinically important and statistically significant improvements in health status that occurred following Synvisc™ injection underscore the true value of the VIP program and establish a link between improved patient outcomes and a newly acquired learner skill when CME professionals apply several essentials of adult learning theory and incorporate a standardized validated health-status measure in design and delivery of a CME enterprise. In an environment of multi-stakeholder demand for evidence of effective use of health care resources, including CME resources, the use of patient outcomes as a measure cannot be ignored. CME professionals need to consider this trend and examine cost-effective ways of incorporating these measures into the design and delivery of future CME endeavors.

Disclosure: Professor Bellamy previously was Professor of Medicine, Epidemiology and Biostatistics at the University of Western Ontario in London, ON; Rory Tekanoff previously was Manager, Continuing Health Education at Rhone-Poulenc Rorer Canada Inc., Montreal, QC; the project was funded by a nonrestricted educational

grant from Rhone-Poulenc Rorer Canada Inc. and Biomatrix Canada.

Acknowledgments

The authors acknowledge the assistance of Jane Campbell for manuscript preparation, Larry Stitt for data analysis, Anna Galteri for data and materials compilation, representatives from Biomatrix (Canada) Inc. and Rhone Poulenc Rorer for logistic support, and to our colleagues in family practice for their enthusiastic participation in the program.

References

1. Badley EM, Lee J. The consultant's role in continuing medical education of general practitioners: the case of rheumatology. BMJ 1987; 294:100–103.
2. Bellamy N. WOMAC osteoarthritis index—user's guide III. London, ON, 1998.
3. Nasmith L, Franco ED. Minor surgical procedures. Faculty development workshop. Can Fam Physician 1997; 43:715–718.
4. Bellamy N, Kirwan J, Boers M, et al. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip and hand osteoarthritis. Consensus development at OMERACT III. J Rheumatol 1997; 24:799–802.
5. Altman R, Brandt K, Hochberg M, et al. Design and conduct of clinical trials in patients with osteoarthritis: recommendations from a task force of the Osteoarthritis Research Society. Osteoarthritis and Cartilege 1996; 4:217–243.
6. SAS Version 6.12. Cary: SAS Institute Inc., Chicago, IL.
7. Boers M, Brooks P, Strand V, Tugwell P. The OMERACT Filter for outcome measures in rheumatology. J Rheumatol 1998; 25:198–199.
8. Glazier RH, Dalby DM, Badley EM, Hawker PGA, Bell MJ, Buchbinder RJ. Determinants of physician confidence in the primary care management of musculoskeletal disorders. Rheumatol 1996; 23:351–356.
9. Davis D. Does CME work? An analysis of the effect of educational activities on physician performance or health care outcomes. Int J Psychiatr Med 1998; 28:21–39.
10. Oxman AD, Thomson MA, Davis DA, Haynes RB. No magic bullets: a systematic review of 102 trials of interventions to improve professional practice. Can Med Assoc J 1995; 153:1423–1431.
11. Davis DA, Thomson MA, Oxman AD, Haynes RB. Changing physician performance. A systematic review of the effect of continuing medical education strategies. JAMA 1995; 274:700–705.
12. Sackett DL, Rosenberg WM. The need for evidence-based medicine. J R Soc Med 1995; 88:620–624.
13. Davis D, Lindsay E, Mazmanian PD. The effectiveness of CME interventions. In Davis DA, Fox RD, eds. The physician as learner. Chicago: American Medical Association, 1994:241–280.
14. Fox, RD. Lessons from the change study: a case for collaborative research in continuing medical education. Teach Learn Med 1990; 2:126–129.
15. Fox RD, Bennett NL. Learning and change: implications for continuing medical education. BMJ 1998; 316:466–468.
16. Pill R, Stott NC, Rollnick SR, Rees M. A randomized controlled trial of an intervention to improve the care given in general practice to Type II diabetic patients: patient outcomes and professional ability to change behaviour. Fam Pract 1998; 15:229–235.
17. Clark NM, Gong M, Schork MA, et al.Impact of education for physicians on patient outcomes. Pediatrics 1998; 101:831–836.
18. Cherkin D, Deyo RA, Berg AO. Evaluation of a physician education intervention to improve primary care for low-back pain. II. Impact on patients. Spine 1991; 16:1173–1178.
19. Rubenstein LV, Calkins DR, Young RT, et al. Improving patient function: a randomized trial of functional disability screening. Ann Intern Med 1989; 111:836–842.
20. Inui TS, Yourtee EL, Williamson JW. Improved outcomes in hypertension after

physician tutorials. A controlled trial. Ann Intern Med 1976; 84:646–651.

21. Fox RD, Mazmanian PE, Putnam RW. Changing and learning in the lives of physicians. New York: Praeger Publications, 1989.
22. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt L. Validation study of WOMAC: a health status instrument for measuring clinically-important patient-relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988; 15:1833–1840.
23. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt L. Validation study of WOMAC: a health status instrument for measuring clinically-important patient-relevant outcomes following total hip or knee arthroplasty in osteoarthritis. J Ortho Rheumatol 1988; 1:95–108.
24. Felson DT, Anderson JJ, Boers M, et al. American College of Rheumatology preliminary definition of improvement in rheumatoid arthritis. Arthritis Rheum 1995; 38:727–735.
25. Wobig M, Dicklut A, Maier R, Vetter G. Viscosupplementation with hylan G-F 20: a 26-week controlled trial of efficacy and safety in the osteoarthritic knee. Clin Ther 1998; 20:410–423.
26. Adams ME, Atkinson MH, Lussier AJ, et al. The role of viscosupplementation with hylan GF-20 (Synvisc) in the treatment of osteoarthritis of the knee: a Canadian multicentre trial comparing hylan GF-20 alone, hylan GF-20 with nonsteroidal anti-inflammatory drugs (NSAIDs) and NSAIDs alone. Osteo Cart 1995; 3:213–226.
27. Lussier A, Cividino A, McFarlane C et al. Viscosupplementation with hylan GF-20 for the treatment of osteoarthritis: findings from clinical practice in Canada. J Rheumatol 1996; 23:1579–1585.

Osteoarthritis and Cartilage (2002) **10**, 506–517

1063–4584/02/$35.00/0
doi:10.1053/joca.2002.0798, available online at http://www.idealibrary.com on IDEAL®

Osteoarthritis and Cartilage | CRS International Cartilage Repair Society | OARSI OSTEOARTHRITIS RESEARCH SOCIETY INTERNATIONAL

A prospective, randomized, pragmatic, health outcomes trial evaluating the incorporation of hylan G-F 20 into the treatment paradigm for patients with knee osteoarthritis (Part 1 of 2): clinical results

J.-P. Raynauld*, G. W. Torrance†‡, P. A. Band§, C. H. Goldsmith†, P. Tugwell‖, V. Walker‡, M. Schultz¶ and N. Bellamy**, in collaboration with the Canadian Knee OA Study Group††

**Department of Medicine, University of Montreal, Montreal, Quebec, Canada*

†*Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada*

‡*Innovus Research Inc., Burlington, Ontario, Canada*

§*Biomatrix, Inc., Ridgefield, New Jersey, U.S.A.*

‖*Department of Medicine, University of Ottawa, and Ottawa Hospital, Ottawa, Ontario, Canada*

¶*Rhône-Poulenc Rorer Canada Inc., St. Laurent, Quebec, Canada*

***Department of Medicine, The University of Queensland, Queensland, Australia*

Summary

Objective: First, to assess the clinical effectiveness of hylan G-F 20 in an appropriate care treatment regimen (as defined by the American College of Rheumatology (ACR) 1995 guidelines) as measured by validated disease-specific outcomes and health-related quality of life endpoints for patients with osteoarthritis (OA) of the knee. Second, to utilize the measures of effectiveness and costs in an economic evaluation (see accompanying manuscript).

Design: A total of 255 patients with OA of the knee were enrolled by rheumatologists or orthopedic surgeons into a prospective, randomized, open-label, 1-year, multi-centred trial, conducted in Canada. Patients were randomized to 'Appropriate care *with* hylan G-F 20' (AC+H) or 'Appropriate care *without* hylan G-F 20' (AC). Data were collected at clinic visits (baseline, 12 months) and by telephone (1, 2, 4, 6, 8, 10, and 12 months).

Results: The AC+H group was superior to the AC group for all primary (% reduction in mean Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain scale: 38% vs 13%, P=0.0001) and secondary effectiveness outcome measures. These differences were all statistically significant and exceeded the 20% difference between groups set *a priori* by the investigators as the minimum clinically important difference. Health-related quality of life improvements in the AC+H group were statistically superior for the WOMAC pain, stiffness and physical function (all P<0.0001), the SF-36 aggregate physical component (P<0.0001) and the Health Utilities Index Mark 3 (HUI3) overall health utility score (P<0.0001). Safety (adverse events and patient global assessments of side effects) differences favoured the AC+H group.

Conclusion: The data presented here indicate that the provision to patients with knee OA of viscosupplementation with hylan G-F 20 within an appropriate care treatment regimen provides benefits in the knee, overall health and health related quality of life at reduced levels of co-therapy and systemic adverse reactions. © 2002 OsteoArthritis Research Society International. Published by Elsevier Science Ltd. All rights reserved.

Key words: Hylan G-F 20, Osteoarthritis, Knee, Effectiveness, Health-related quality of life, Randomized controlled trial.

Abbreviations: ACR, American College of Rheumatology; OA, osteoarthritis; AC+H, Appropriate care *with* hylan G-F 20; AC, Appropriate care *without* hylan G-F 20; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index; HUI3, Health Utilities Index; NSAIDS, nonsteroidal anti-inflammatory drugs; FDA, Food and Drug Administration; RCT, randomized controlled trial; CRO, contract research organization; HRQOL, health-related quality of life; SF-36, Short Form 36; ITT, intent-to-treat; ANCOVA, analysis of covariance; GI, gastrointestinal.

Received 4 March 2001; revision requested 10 October 2001; revision received 19 November 2001; accepted 4 February 2002.

††The Canadian Knee OA Study Group included the Steering Committee and clinical investigators. The Steering Committee consisted of J. P. Raynauld and G. W. Torrance, co-principal investigators; N. Bellamy, C. H. Goldsmith, and P. Tugwell, co-investigators; P. A. Band, M. Schultz, and V. Walker. The clinical investigators were: M. J. Bell, MD, Department of Medicine, University of Toronto, Toronto, ON; R. B. Bourne, MD, Department of Surgery, University of Western Ontario, London, ON; R Cloutier, MD, Departement D'Orthopedie, Centre Hospitalier Affillie, Quebec, QC; A. M. Crawford, MB, Department of Medicine, University of Calgary, Calgary, AB; J. G. Cinats, Department of Surgery, University of Alberta Hospitals, Capital Health Regional Health Authority, Edmonton, AB; M.-A. Fitzcharles, MD, Department of Medicine, McGill University, Montreal, QC; R. D. Inman, MD, Division of Rheumatology, Toronto Western Hospital, University of Toronto, Toronto, ON; K. W. Marshall, MD, Division of Orthopaedics, Toronto Western Hospital, University of Toronto, Toronto, ON; D. P. O'Hanlon, MB, Department of Rheumatology, University of British Columbia, North Vancouver, BC; W. P. Olszynski, MD, Department of Medicine, University of Saskatchewan, Saskatoon, SK; J. Pope, MD, Department of Medicine, University of Western Ontario, London, ON; J.-P. Raynauld, MD, Department of Medicine, University of Montreal, QC; E. D. Sutton, MD, Department of Medicine, Dalhousie University, Halifax, NS; J. C. Thorne, MD, Department of Internal Medicine and Rheumatology, York County Hospital, Newmarket ON.

The study was funded by Biomatrix, Inc and Rhône-Poulenc Rorer Canada Inc.

Address correspondence to: Dr Jean-Pierre Raynauld, Montreal Institute of Rheumatology, 1551 Ontario Street East, Montreal, PQ Canada H2L 1S6. Tel: (514) 523-5273, Fax: (514) 523-5973, E-mail: jp.raynauld@videotron.ca

Introduction

Osteoarthritis (OA) is a common, degenerative musculoskeletal condition which consistently challenges the practising clinician and adds substantial burden to health care budgets[1,2]. The increased prevalence of OA with aging, coupled to the demographics of aging populations, make OA a high priority health care problem[3]. OA is a leading cause of severe activity limitations and disability, with indirect costs to society, which can far exceed its direct medical costs[4].

Guidelines for managing the symptoms of knee OA are available from various sources[5]. The goal of therapy is to control pain and maintain function. Weight control, physical therapy and simple analgesics such as acetaminophen, are suggested as first-line tools for patient management to minimize the need for higher risk treatments such as non-steroidal antiinflammatory drugs (NSAIDs) or surgery. NSAIDs continue to be a widely used medical therapy in response to patient demands for symptomatic improvement. In the United States alone, there are an estimated 56 000 hospitalizations and 8600 deaths each year among OA patients, attributed to NSAID treatment[6].

Viscosupplementation is a new therapy for the treatment of knee OA based on the replacement of synovial fluid by intraarticular injection of viscoelastic solutions containing hyaluronan or its derivatives[7]. Hylan G-F 20 (Synvisc® Genzyme Corporation, Cambridge MA U.S.A.) is one of the viscosupplementation products approved for marketing in Canada since 1992 and the United States since 1997 after public review of the data by a Food and Drug Administration (FDA) advisory panel[8]. A recent systematic review of the randomized controlled trial (RCT) data on viscosupplementation concluded, that despite mixed results, the overall data support the efficacy of viscosupplementation[9]. While some physicians continue to question the efficacy of hylan G-F 20, the reality is that hylan G-F 20 is an approved treatment in Canada, the U.S.A., and most other countries. Furthermore, the recently revised guidelines published by the American College of Rheumatology (ACR) now include viscosupplementation in the treatment paradigm for knee OA, thus establishing it as a standard therapy[10].

Considering the limited resources available for health care, it is important to consider how incorporating the new technology affects patient outcomes and health care expenditures. A randomized, controlled trial of health outcomes was specifically designed to determine the *incremental* effectiveness, cost-effectiveness and cost-utility of making viscosupplementation with hylan G-F 20 available as part of an appropriate care paradigm for treating patients with knee OA. The study utilized a pragmatic design to maintain a real world scenario, and therefore measured effectiveness rather than efficacy[11–14]. That is, rather than asking the question of whether the treatment is efficacious compared to placebo, the trial sought to determine whether the treatment was effective under real world conditions. The Canadian Guidelines for Economic Evaluation of Pharmaceuticals state: 'Ideally, pharmacoeconomic studies should report on drug effectiveness rather than efficacy'[15]. For this reason the trial design minimized protocol-driven interventions and the comparator arm did not include placebo injections. Effectiveness includes all aspects of a treatment that add or detract from its success, including efficacy, patient compliance and satisfaction, safety, and positive or negative interactions with other concurrent treatments.

The availability of a safe and effective local therapy for managing a localized condition such as knee OA might offer important health care benefits. The clinical results and health-related quality of life (HRQOL) outcomes for this trial are reported here, with the economic results separately reported in an accompanying manuscript[16].

Materials and methods

STUDY MANAGEMENT

The study was funded jointly by Biomatrix, Inc and Rhône-Poulenc Rorer Canada Inc. Innovus Research Inc., an independent contract research organization (CRO), was contracted to manage the study. An independent Steering Committee was assembled with the responsibility to design the study, develop the analysis plan, resolve methodological issues that arose throughout the study, and interpret and disseminate study results. The Committee consisted of five academics, one representative from each of the two sponsoring companies and one representative from the CRO. The Steering Committee was deliberately structured to be dominated by the five independent academics on the Committee. The Steering Committee actively dealt with all scientific questions that arose throughout the course of the study, and did so blinded to implications. The contractual arrangement gave the investigators unrestricted rights to publish the study results.

PATIENTS

Patients were enrolled between April and December 1997, at 14 sites across Canada (10 rheumatologists, four orthopedic surgeons). The study protocol and informed consent form were approved by the relevant Ethics Committees for the sites. Informed consent was obtained from each patient.

Patients with age greater than 40 years, were required to have a primary diagnosis of radiologically verified OA in the study knee (knee most symptomatic or with the most predominant musculoskeletal problem), to be symptomatic [visual analogue scale total pain score greater than 175 mm of 500 mm on the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain scale] despite prior treatment with acetaminophen or NSAIDS at any point prior to the study, to be ambulatory and willing to participate and sign informed consent. Patients with Grade IV[17] radiologic changes according to the clinical investigators were excluded. Other exclusion criteria included patients with inflammatory arthropathies, a tense effusion in the study knee at baseline, chondrocalcinosis or those with a severe varus or valgus deformity in the study knee. In addition, patients were excluded if they had received a steroid injection in the study knee during the prior 3 months, if they had prior viscosuppiementation therapy, if they had isolated patellofemoral OA or any uncontrolled morbidity, particularly morbidity in any joint which would impede measurements in the study knee.

TRIAL DESIGN

This was a multicentre, 1-year, prospective, randomized, open-label study. Patients were randomized to either 'appropriate care *with* hylan G-F 20' (AC+H) or to 'appropriate care *without* hylan G-F 20' (AC). The AC group only

differed in that intraarticular injections of hylan G-F 20 or other viscosupplementation products were not allowed. Appropriate care was the preferred management strategy of a treating physician who was encouraged to follow the Guidelines for the Medical Management of Osteoarthritis of the Knee proposed by the ACR[5]. Appropriate care could include medications such as analgesics, NSAIDs, corticosteroid injections, supportive measures such as education and counseling, weight loss, joint rest, application of heat or ice, and use of devices, physical therapy, arthroscopy, and total joint replacement.

Hylan G-F 20 is administered as a series of 3 intra-articular injections at intervals of 1 week. The contralateral knee could also be treated with hylan G-F 20, and patients could receive subsequent treatments to either or both knees as required.

Computer-generated randomization was designed to be balanced (1:1 allocation ratio) within each site. Randomization within site was blocked, but the block size was randomly assigned as blocks of 2 or 4, with the additional constraint that blocking was balanced for the first 12 patients. Additional patients exceeding the first block of 12, were randomly allocated in blocks of 2. The allocation scheme was concealed from all clinical sites. Central randomization was used whereby the site telephoned the CRO, provided the patient's initials, and received the patient's identification (ID) number and treatment allocation.

Patients were assessed at the site during the baseline visit and the 12-month termination visit. Patients randomized to AC+H returned to the site for 2 consecutive weeks after baseline for the remaining hylan G-F 20 injections. These were the only site visits required by the protocol. Structured telephone interviews of the patients in both treatment groups were conducted by the CRO at months 1, 2, 4, 6, 8, 10, and 12. The 12-month termination visit was included for patient assessment by the investigator and for measuring change since baseline. Patients returned to the physician as required for clinical deterioration, treatment of adverse events, change in medication, or additional treatment with hylan G-F 20 if required.

Patient demographics, appropriate care treatment for knee OA, treatment for overall OA, concomitant medications, and patient self-administered questionnaires were collected at the baseline visit. The same information was collected during the telephone interviews, with the addition of pill counts performed by the patient, medication dosage and duration, adverse events, health care resources, and whether the health care resource was related to OA. The patients kept a diary to keep track of this information, and their content was provided to the telephone interviewer at each telephone interview. During the telephone interviews, the patient referred to the self-administered questionnaire and provided his/her answers to the telephone interviewer. To blind the patient to his/her previous answers to the same questions, s/he was instructed not to record the answers, and the questionnaire was laminated with plastic to make it difficult if someone tried to do so. Although the questionnaires were completed originally at the baseline visit and then during telephone interviews, a study comparing the completion of the WOMAC Likert 3.0 questionnaire at the physician's office to completion over the telephone found that differences between the modes of administration did not reach statistical significance[18]. Information collected during the telephone interviews (with the exception of the questionnaires) was compared to the patient's medical chart during monitoring visits and differences were resolved. The investigator reviewed the adverse events for possible attribution to study interventions.

OUTCOMES

The primary measure of effectiveness was the mean change in the WOMAC Likert 3.0 pain score in the study knee from baseline to termination. The WOMAC is a self-administered disease-specific HRQOL instrument that asked the patient questions concerning his/her study knee[19]. The WOMAC Likert 3.0 provides scores for three subscales: pain, stiffness, and physical functioning, and an aggregate total score.

There were also measures of secondary effectiveness. Two of the secondary effectiveness measures were the percent of patients improved at termination since baseline using different combinations of the WOMAC Likert 3.0 subscales as follows: (1) at least 20% improvement since baseline in the WOMAC pain score in the study knee; (2) at least 20% improvement since baseline in the WOMAC pain score in the study knee and *either* 20% improvement in function score or stiffness score. A 20% difference between treatment groups for the primary and secondary measures of effectiveness was established a priori by the Steering Committee as the minimum clinically important difference, in part based on previous research[20]. Other secondary effectiveness measures were the patient global assessment of effectiveness for (1) OA in study knee; (2) OA in all joints, and (3) overall health.

HRQOL was measured using three instruments: disease-specific HRQOL using the WOMAC; general HRQOL using the Short Form 36 (SF-36)[21], and preference-based HRQOL using the Health Utilities Index Mark 3 (HUI3)[22]. The SF-36 provides two composite scales: aggregate physical component and aggregate mental component. The HUI3 provides an overall multi-attribute utility score (min: −0.36, death: 0, max: 1). The overall utility score is the preference or worth assigned to a particular health status on an interval scale where 0 represents death and 1 represents perfect health. States worse than death can take on negative scores.

Safety was measured in two ways during the course of the study. The first method was by asking patients to report adverse events during each telephone interview and then having the clinical site review the adverse events. The second method of measuring safety was by asking patients to complete global assessments of side effects. Global assessments were measured in two ways: throughout the study at baseline and at each telephone interview recalling the past 4 weeks; and once during the 12-month termination visit recalling the time period since the baseline visit.

STATISTICS

The sample size was calculated to detect a 20% difference between treatment groups in the primary effectiveness measure. Using a power of 90% and α=0.05, the required sample size was 94 patients per group, for a total of 188 patients. The final total sample size required was 252 patients, to accommodate a 20% predicted dropout rate over 1 year and to accommodate stratification by site (15 sites).

All patients enrolled in the study were included in the intent-to-treat (ITT) group for all analyses. However, if a patient in the AC group violated the protocol by receiving hylan G-F 20 treatment, the patient was treated as a

dropout at that point, and all data collected after that time were not included in the analyses. These patients were classified as crossovers, and their data following the hylan G-F 20 treatment were imputed as was done for all dropouts. This was necessary to ensure the analysis was consistent with a comparison of appropriate care in a world with hylan G-F 20 to appropriate care in a world without hylan G-F 20.

Two models were used for the statistical analyses, and results for the first model are provided. The first model adjusted for design variables (baseline value of the variable being analysed, site, blocking by site, BMI, Baseline WOMAC aggregate score), and the second model adjusted for design variables and potentially clinically important differences (as judged by the clinical principal investigator while blinded to treatment allocation) between the treatment groups at baseline.

An analysis of covariance (ANCOVA) was used for the primary effectiveness analysis and the HRQOL analysis. A generalized linear model was performed for analysis of patients improved. A logistic analysis was undertaken for the patient global assessment of side effects and effectiveness. A nested analysis that incorporates the number of events per patient was used to compare the number of gastrointestinal (GI) adverse events.

The hot deck method[23] was utilized to impute data for the primary and secondary effectiveness of patient improved. Dropout patients were matched with a patient who completed the study. The matched patient was randomly selected from the group of patients who matched the dropout patient on criteria deemed most relevant in predicting primary effectiveness. The Last Observation Carried Forward (LOCF) imputation technique was performed to compare to the hot deck method.

Results

PATIENT CHARACTERISTICS

A total of 255 patients were enrolled, 127 patients randomized to AC+H and 128 to AC (Fig. 1). The central randomization process was audited to ensure that the randomization schedule was implemented properly. There were more dropout patients in the AC group (21) than the AC+H group (3) (P=0.001). Of the 21 patients who dropped out of the AC group, the main two reasons were that the patients wanted hylan G-F 20 (eight patients) and that the patients were unwilling to continue (eight patients). As shown in Fig. 1, eight patients randomized to AC received hylan G-F 20 (protocol violators/crossovers), and one patient in the AC+H group did not receive hylan G-F 20 (protocol violator/crossover). The patient changed their mind after being randomized to receive hylan G-F 20. Eighteen of the 24 patients (75%) who dropped out, did so before Month 4. Of the 24 patients who dropped out, four continued to have data collected during the remainder of the study, however, the data were not used in the analyses. Because these four patients violated the protocol by receiving hylan G-F 20 treatment despite being randomized to AC without hylan G-F 20, their data after the hylan G-F 20 injection were not included in the analyses.

Demographic and OA status data are presented in Table I. Greater than 79% of patients in both groups had received previous acetaminophen and NSAID treatment for OA in their knee(s). Although Grade IV OA in the study knee as determined at the sites by the investigators at enrollment was an exclusion criterion, 20% of patients in the AC+H group and 33% of patients in AC had grade IV OA as subsequently determined by central radiologic grading. Greater than 84% of patients in both groups had OA in the other knee, and greater than 68% of patients in both groups had other joints affected.

KNEE OA TREATMENT

Table II lists knee OA and overall OA treatment. All patients except one in the AC+H group had at least one course of hylan G-F 20 in their study knee, and 53 (42%) had at least 1 course in their other knee. Forty-eight patients (38%) in the AC+H group received a second course in the study knee, three patients (2%) received a third course in their study knee, and 20 patients (16%) received a second course in their other knee (data not shown in Table II). There were more patients in the AC group who reported corticosteroid injection(s) in the study knee (89 vs 18) or the other knee (35 vs 8) (both P<0.0001). There were more corticosteroid injections in the AC group in the study knee (149 vs 27) and the other knee (51 vs 14). There were more patients in the AC group taking NSAIDs for any knee (P=0.0062), and other medications for any knee (P=0.0216). Other medications included medications such as antiinflammatories, neuralgia therapy, opioid analgesics and vitamins. There were seven arthroscopies and four total knee replacements in the AC group compared to one arthroscopy and two total knee replacements in the AC+H group. Despite these reductions in the use of medication for the study knee, there was no significant difference between the groups in the utilization of concomitant medications for overall OA (Table II).

The other treatments, not listed in Table II, that were used most often in both groups were exercise, physiotherapy, walking, water exercises, and assistive devices such as bandages, canes, knee braces, bath bars, and orthotics. There were too many details to provide the other treatments and assistive devices results in Table II. However, the cost results summarized in the accompanying economic manuscript indicate that the annual cost per patient for other therapy was $5 in the AC+H group versus $16 in the AC group. The annual cost per patient for assistive devices was $237 in the AC+H group versus $305 in the AC group[16].

EFFECTIVENESS

Table III provides the primary and secondary effectiveness results. The AC+H group was superior to the AC group for all primary and secondary effectiveness measures. These differences were all statistically significant and exceeded the 20% minimum clinically important difference. The AC+H group experienced a 25% greater improvement in the WOMAC pain score in the study knee from baseline to termination (P=0.0001). The AC+H group had a larger percent of patients who improved by at least 20% (P=0.0001). The primary and secondary effectiveness analyses yielded similar results for model 2 (data not shown). Imputation using LOCF did not change the results. The AC+H group experienced 26% greater improvement in the WOMAC pain score, and a 30% greater improvement in percent of patients who improved by at least 20% in WOMAC pain. The AC+H group also did better on the patient global assessments of effectiveness for OA in the

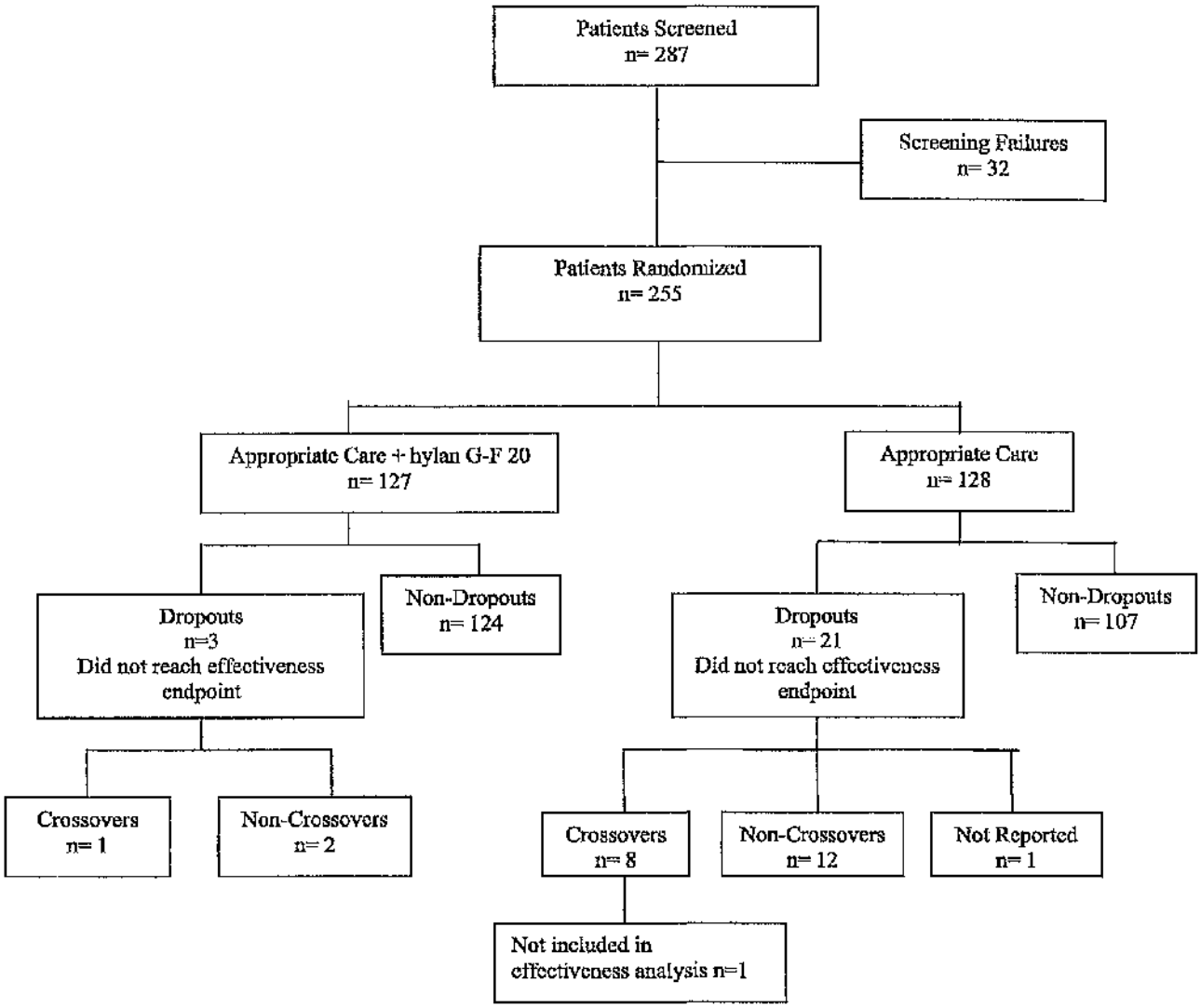


Fig. 1. Trial profile.

study knee, as well as OA in all joints and overall health ($P<0.05$).

Figure 2 displays the mean WOMAC pain score at each time point during the study year. The patients in the AC+H group had a greater reduction in the WOMAC pain score over the full study year compared to the AC group ($P=0.0001$).

HEALTH-RELATED QUALITY OF LIFE

Table IV provides the baseline, Month 12, change, and change as a % of baseline for the three HRQOL outcome measures. For all three WOMAC subscales, the SF-36 aggregate physical component, and the HUI3 overall health utility score, the AC+H group was statistically significantly superior ($P<0.0001$). For all cases in the WOMAC and SF-36, the difference between groups was greater than 20% except for the SF-36 aggregate physical component where the difference between groups was 19%.

SAFETY AND TOLERABILITY

Adverse events were reported by 96% of patients (1114 events) in the AC+H group and 90% of patients (1026) in the AC group (not compared statistically). There was one serious adverse event in the AC group (patient presented to the emergency room with a gastro-duodenal ulcer) listed by the investigator as remotely related to appropriate care.

Intraarticular injection of hylan G-F 20 is occasionally accompanied by pain, swelling, or effusion in the treated knee. A local adverse event was defined during the analysis as any emergent signs or symptoms occurring in the knee. The local adverse events were subdivided into those occurring within 48 hours of a hylan G-F 20 injection and those occurring at any other time. There were 82 local adverse events (in 38 patients) that occurred within 48 hours of a hylan G-F 20 injection in the AC+H group. Of these 82 local adverse events, one was reported as related to osteoarthritis, nine were reported as not related to hylan G-F 20, 15

Table I
Demographic information and osteoarthritis status

Demographics, f (percent of *n*)*	AC+H (*n*=127)	AC (*n*=128)
Age in years, mean (s.d.)	62.6 (9.4)	63.5 (10.5)
Sex		
Female	86 (68%)	93 (73%)
Body mass index (kg/m^2), mean (s.d.)	32.1 (6.0)	32.9 (7.2)
OA status		
Duration (years) of OA symptoms		
Study knee, mean (s.d.)	9.0 (9.5)	9.9 (9.7)
Other knee, mean (s.d.)	7.4 (8.8)	8.3 (9.3)
Previous therapy for OA of the knee(s)		
Acetaminophen	100 (79%)	109 (85%)
NSAIDs	120 (94%)	110 (86%)
Prior surgery, study knee	40 (31%)	39 (30%)
Prior surgery, other knee	27 (21%)	23 (18%)
Radiology grading within 1 year (central grading)†		
Not reported	0 (0%)	1 (1%)
Grade 0	4 (3%)	4 (3%)
Grade I	17 (13%)	11 (9%)
Grade II	32 (25%)	33 (26%)
Grade III	49 (39%)	37 (29%)
Grade IV	25 (20%)	42 (33%)
OA at baseline		
Other knee affected	109 (86%)	108 (84%)
Any other joints affected	95 (75%)	87 (68%)
Patient global assessment of OA in study knee at baseline		
Not reported	0 (0%)	1 (1%)
Very good	0 (0%)	0 (0%)
Good	2 (2%)	1 (1%)
Fair	44 (35%)	31 (24%)
Poor	58 (46%)	57 (45%)
Very poor	23 (18%)	38 (30%)
Patients global assessment of OA in all joints at baseline		
Not reported	1 (1%)	2 (2%)
Very good	1 (1%)	2 (2%)
Good	9 (7%)	5 (4%)
Fair	54 (43%)	44 (34%)
Poor	47 (37%)	49 (38%)
Very poor	15 (12%)	26 (20%)
WOMAC pain subscale score (0–20), mean (s.d.)	11.4 (2.7)	11.9 (2.9)

*f is frequency, *n* is sample size. Not all percentages sum to 100 due to rounding.
†Radiology grading is based on central grading, which may have differed from the site investigator's determination for patient eligibility.
OA=osteoarthritis; WOMAC=Western Ontario and McMaster Universities Osteoarthritis Index; AC+H=Appropriate Care+hylan G-F 20; AC=Appropriate Care.

were reported as remotely, possibly or probably related to hylan G-F 20, and 57 related to the injection procedure.

The occurrence of GI adverse events was lower in the AC+H group for total GI events (109 vs 140 events, P=0.0439), and GI events attributed to AC (25 vs 62 events, P=0.0001), and for total severe GI events (26 vs 53, P=0.0033), and severe GI events attributed to AC (5 vs 22, P=0.0024). Medications taken for side effects of OA treatment were collected. Thirty-nine patients in the AC group were taking medications for the gastrointestinal tract compared to 21 patients in the AC+H group (P=0.0057).

For the global assessments of side effects for the time period since baseline, 62% (79/127) of AC+H patients experienced no side effects compared to 41% (52/128) of AC patients (P=0.0100). The global assessments of side effects (combined mild, moderate or severe) performed at baseline and months 1, 2, 4, 6, 8, 10 and 12 are illustrated in Fig. 3. Fewer patients in the AC+H group (52%; 64/124) experienced side effects at Month 12 than patients in the AC group (68%; 73/107) (P=0.0116).

Discussion

This report details the clinical results of a prospective, randomized, effectiveness/health outcomes trial evaluating the incremental value of making a new treatment modality, viscosupplementation with hylan G-F 20, available for the treatment of patients with knee OA. All of the clinical outcomes measured provided consistent results favoring the group receiving AC+H. The difference between the groups was clinically important and statistically significant using a disease-specific instrument (WOMAC 3.0), a generic HRQOL instrument (SF-36), a preference based HRQOL instrument (HUI3) and global evaluations by the patient of OA in the study knee, overall OA, and overall

Table II
Knee osteoarthritis treatment and overall osteoarthritis treatment

Treatment, f (percent of *n*)*	AC+H (*n*=127)	AC (*n*=128)	*P*-value
Number of patients reporting hylan G-F 20 course(s)			
Study knee	126 (99%)	6 (5%)	
Other knee	53 (42%)	0 (0%)	
Number of patients reporting corticosteroid injection(s)			
Study knee	18 (14%)	89 (70%)	<0.0001
Other knee	8 (6%)	35 (27%)	<0.0001
Number of patients reporting arthroscopy			
Study knee	1 (1%)	5 (4%)	
Other knee	0 (0%)	2 (2%)	
Number of patients reporting total knee replacement (TKR)			
Study knee	1 (1%)	3 (2%)	
Other knee	1 (1%)	1 (1%)	
Number of patients reporting medication for any knee			
Analgesic (oral)	84 (66%)	96 (75%)	0.1158
NSAID (oral)	82 (65%)	101 (79%)	0.0062
Alternative therapy	41 (32%)	44 (34%)	0.5501
Analgesic (topical)	20 (16%)	24 (19%)	0.6996
Other	13 (10%)	25 (20%)	0.0216
Number of patients reporting medications for overall osteoarthritis‡			
Musculoskeletal	16 (13%)	15 (12%)	
CNS	15 (12%)	12 (9%)	
Minerals and vitamins	1 (1%)	1 (1%)	
Anti-infectives	1 (1%)	0 (0%)	
Other	5 (4%)	10 (8%)	

*f is frequency, *n* is sample size.
†*P*-value results from Model 1 adjusting for design variables.
‡If a patient was taking the same medication for knee osteoarthritis and osteoarthritis in other joints, it was included in knee osteoarthritis.
NSAIDs=non-steroidal antiinflammatory drugs; CNS=central nervous system; AC+H=Appropriate Care+hylan G-F 20; AC=Appropriate Care.

Table III
Primary effectiveness and secondary effectiveness results

	AC+H (*n*=127)	AC (*n*=128)	[(AC+H)−(AC)]	*P*-value*
Primary effectiveness	*n*=127	*n*=127		
Change from baseline to termination in WOMAC pain, mean (s.d.)	−4.4 (3.9)	−1.8 (3.8)	−2.6	0.0001
Change as a % of baseline, mean (s.d.)	−38.4 (34.4)	−13.3 (39.9)	−25.07	<0.0001
Secondary effectiveness f (percent of *n*)†	*n*=127	*n*=127		
Patients improved at termination since baseline:				
WOMAC pain	87 (69%)	51 (40%)	29%	0.0001
WOMAC pain and either stiffness or physical functioning	79 (62%)	46 (36%)	27%	0.0001
Patients global assessment of change since baseline (improved slightly, moderately, or markedly):				
OA in study knee	93 (73%)	35 (27%)	46%	<0.0001
OA in all joints	48 (38%)	22 (17%)	21%	0.0011
Overall health	48 (38%)	21 (16%)	22%	0.0010
Patients global assessment at month 12 over the past 4 weeks (fair, good, or very good):	*n*=124	*n*=107		
OA in study knee	94 (76%)	46 (43%)	33%	<0.0001
OA in all joints	88 (71%)	45 (42%)	29%	<0.0001
Overall health	118 (95%)	91 (85%)	10%	0.0115

*P-value results from Model 1 adjusting for design variables. The results were similar for model 2 adjusting for design variables and potentially clinically important differences at baseline.
†f is frequency, *n* is sample size. The sample size is indicated in the table heading unless otherwise indicated in the table.
OA=osteoarthritis; WOMAC=Western Ontario and McMaster Universities Osteoarthritis Index; AC+H=Appropriate Care+hylan G-F 20; AC=Appropriate Care.

health. These data do not address the continuing debate regarding the relative contribution of the intraarticular procedure and the material injected into the knee. However, they clearly demonstrate that making viscosupplementation available as part of an AC treatment regimen results in clinically important improvement to patients with knee OA.

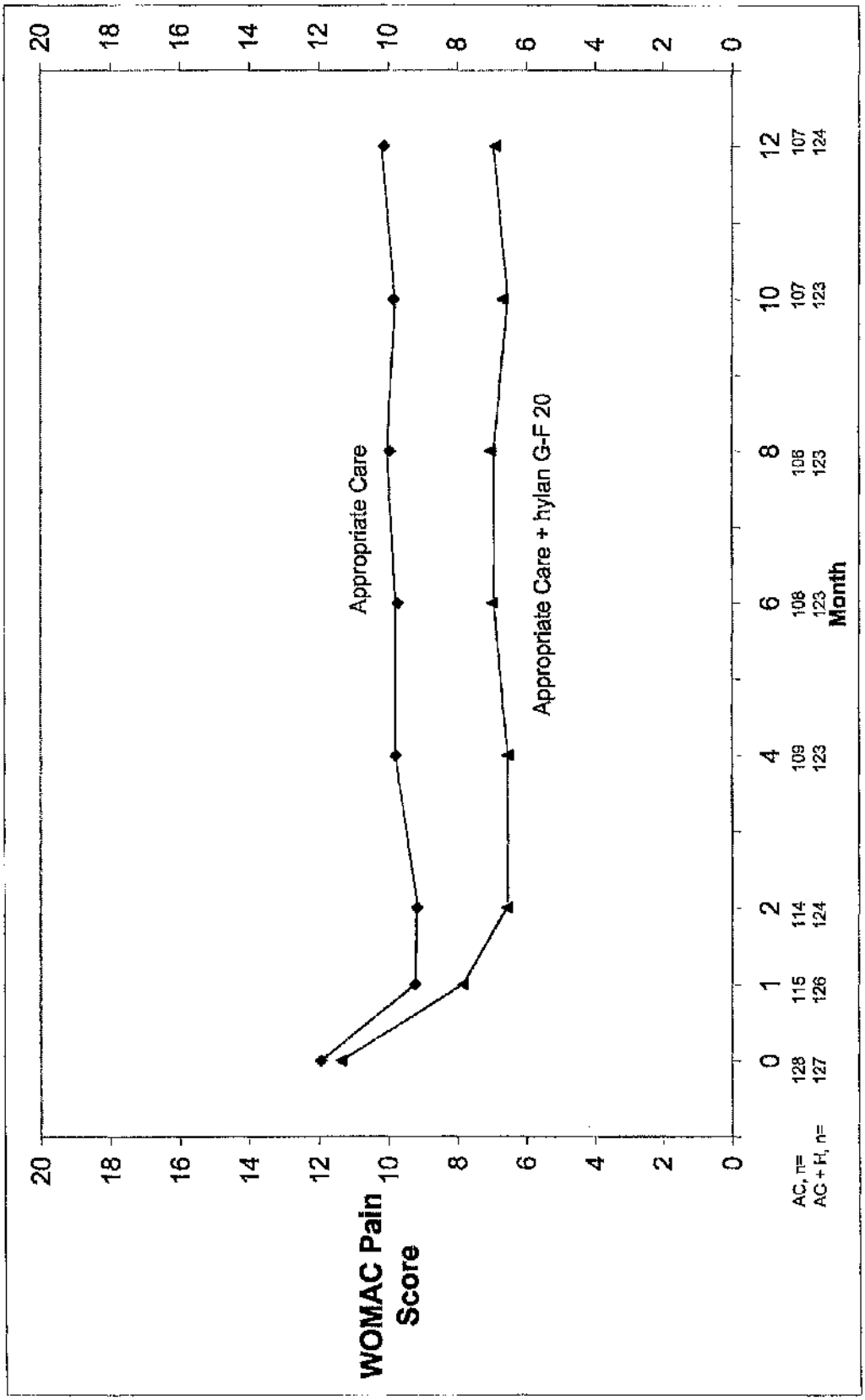


Fig. 2. WOMAC mean pain subscale score by month. WOMAC=Western Ontario and McMaster Universities Osteoarthritis Index.

Table IV
Mean change since baseline in WOMAC, SF-36 and HUI

	Baseline	Month 12	Change (month 12–baseline)*	Change as a % of baseline	*P*-value† for comparison at month 12
WOMAC‡ subscales, mean (s.d.)					
Pain (min: 0; max: 20)					
AC+H	*n*=127	*n*=127			
	11.35 (2.71)	6.94 (3.97)	−4.41 (3.88)	−38.41 (34.39)	<0.0001
AC	*n*=127	*n*=127			
	11.94 (2.89)	10.10 (4.24)	−1.83 (3.83)	−13.34 (39.86)	
(AC+H)−(AC)				−25.07	
Stiffness (min: 0; max: 8)					
AC+H	*n*=127	*n*=124			
	5.06 (1.51)	3.22 (1.74)	−1.83 (1.73)	−34.74 (35.00)	<0.0001
AC	*n*=127	*n*=107			
	5.10 (1.42)	4.31 (1.56)	−0.71 (1.57)	−10.42 (37.42)	
(AC+H)−(AC)				−24.32	
Physical function (min: 0; max: 68)					
AC+H	*n*=127	*n*=124			
	39.54 (9.27)	24.26 (12.95)	−15.04 (12.29)	−37.82 (31.44)	<0.0001
AC	*n*=127	*n*=107			
	40.20 (9.26)	33.87 (13.88)	−5.85 (11.18)	−14.52 (30.39)	
(AC+H)−(AC)				−23.30	
SF-36§, mean (s.d.)					
Aggregate physical component (min: 2; max: 76)					
AC+H	*n*=127	*n*=124			
	28.33 (6.60)	33.24 (10.16)	4.88 (9.78)	20.31 (37.43)	<0.0001
AC	*n*=126	*n*=107			
	28.18 (7.78)	27.78 (8.90)	−0.40 (7.22)	1.07 (29.10)	
(AC+H)−(AC)				19.24	
Aggregate mental component (min: −2; max: 61)					
AC+H	*n*=127	*n*=124			
	51.74 (11.83)	55.29 (10.45)	3.32 (12.06)	11.53 (33.80)	0.0939
AC	*n*=126	*n*=107			
	49.91 (11.82)	52.65 (11.56)	1.55 (10.55)	5.40 (23.11)	
(AC+H)−(AC)				6.13	
HUI3‖ (min: −0.36; max 1), mean (s.d.)					
AC+H	*n*=123	*n*=122			
	0.50 (0.22)	0.63 (0.25)	0.13 (0.23)	n/a	<0.0001
AC	*n*=126	*n*=107			
	0.46 (0.24)	0.51 (0.28)	0.03 (0.22)	n/a	
(AC+H) – (AC)			0.10		

*Due to differences in sample size from baseline to month 12 computation of change (Month 12−Baseline) was calculated for patients with both baseline and termination values.

†*P*-value results from Model 1 adjusting for design variables. The results were similar for model 2 adjusting for design variables and potentially clinically important differences at baseline

‡The higher the score, the worse the problem

§The higher the score, the better the health perception

‖The higher the score, the better the overall health utility

n/a denotes not applicable. Because HUI3 is an interval scale, percent improvements are not useful and indeed distort the magnitude of change.

WOMAC=Western Ontario and McMaster Universities Osteoarthritis Index; AC+H=Appropriate Care+hylan G-F 20; AC=Appropriate Care; SF-36=Short Form 36; HUI3=Health Utilities Index 3.

The data reported that inclusion of hylan G-F 20 in an appropriate care treatment regimen resulted in a meaningful decrease in the utilization of other treatments for knee OA. These decreases were statistically significant with respect to the utilization of steroid injections, oral NSAID therapy and 'other' medications for knee OA. Patients in the AC+H group also received fewer arthroscopies and fewer total knee replacements, but the difference was not compared statistically.

Overall the safety data collected and analysed in this trial confirm that patients treated in different ways are likely to experience different patterns of side effects. Patients in the AC+H group experienced some discomfort associated with the intraarticular procedure. The 15 local adverse events categorized as remotely, possibly or probably attributed to hylan G-F 20 out of a total of approximately 700 hylan G-F 20 injections represents a rate of approximately 2%, similar to that observed in other trials[24]. However, the hylan G-F 20 treated patients also had a clinically meaningful decrease in both the number and severity of GI side effects related to appropriate care and the need for medication to treat GI side effects. Furthermore based on the patients'

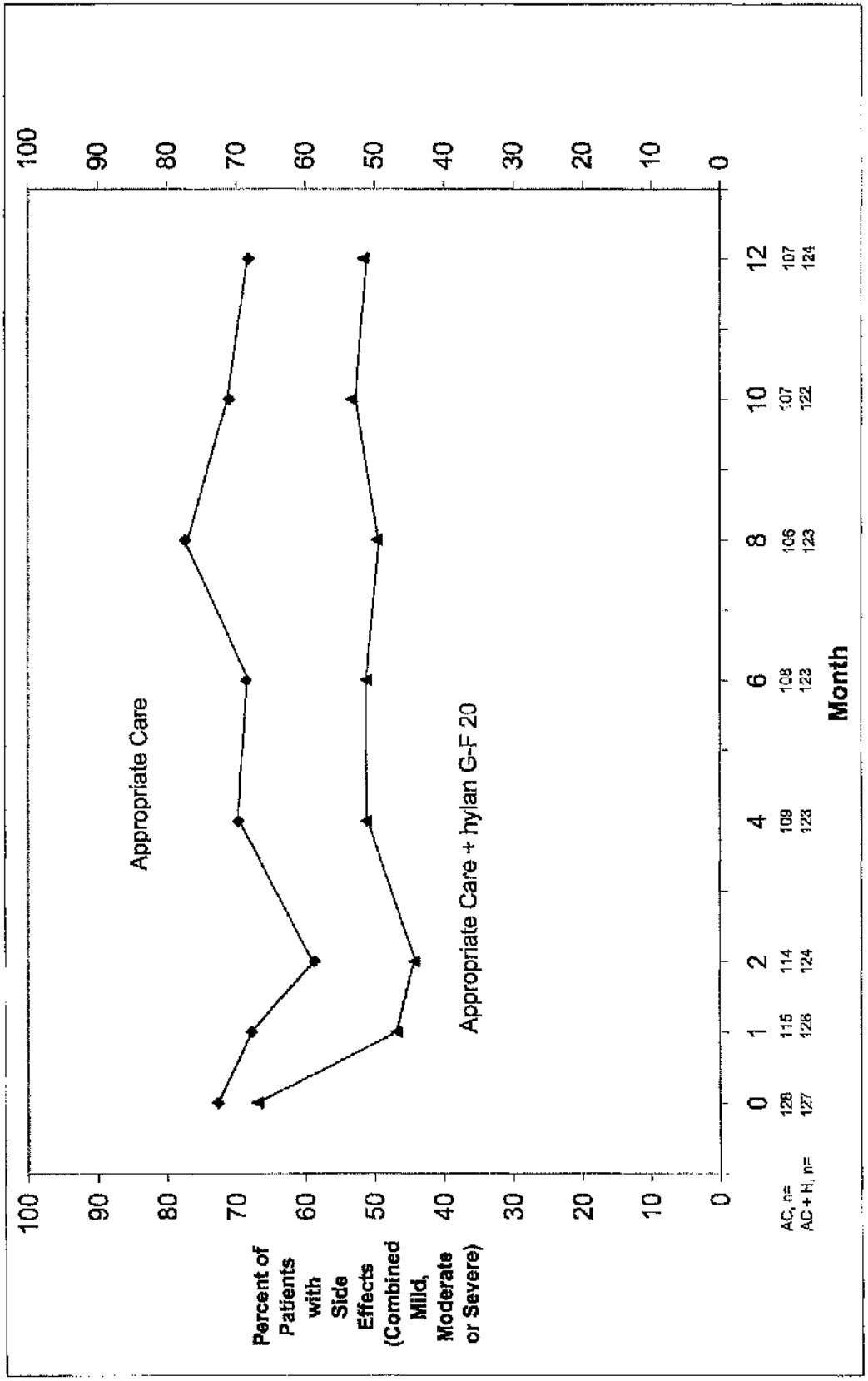


Fig. 3. Percent of patients with global assessment of side effects (combined mild, moderate, or severe) by month.

global evaluation of side effects, it would appear that the patients in the AC+H group judged themselves to have experienced additional benefits by virtue of having encountered fewer side effects. These data suggest that a management strategy, which includes hylan G-F 20 may result in important safety gains, principally by reducing GI events and the necessity for their treatment with GI medications. It should be noted that COX-2 selective inhibitors were not available during the trial.

Although viscosupplementation with hylan G-F 20 is a local treatment which was only used to treat knee OA in this trial, the AC+H group experienced significantly better improvements in global evaluations measuring overall OA, and overall health, and in HRQOL instruments which reflect the health status of the whole patient. This is particularly surprising considering that in the AC+H group 68% of patients had OA in some joint other than the knee and 49% of patients scored their OA in all joints as poor or very poor at baseline (Table I). These 'whole patient' improvements probably reflect the fact that for the patients in this trial the knee was their most symptomatic musculoskeletal problem, and was therefore a major determinant of their pain, disability and HRQOL. It is not uncommon in OA patients for one or two joints to be the primary source of the patient's disability[25]. Similar improvements in HRQOL are observed after surgical treatments for knee OA such as knee replacement[26].

In keeping with the study's pragmatic design, the X-ray grade used to determine study inclusion was that scored by the investigator entering the patient, and based on their best clinical judgement and the radiologist report. The investigator was not asked to provide a grade, but to determine that the patient had OA that was not Grade IV. Hence one radiologist will possibly provide a different rating than the impression of 14 investigators who were not asked to provide a grade. Because the authors were sensitive to potential differences between investigators and to the prevalence of 'borderline' scores, central grading was performed by a trained radiologist. This was done after the patients were entered and used only in the analyses. Patients judged to have Grade IV X-ray by central scoring were not asked to leave the trial.

Grade IV OA of the study knee was an exclusion criterion, because those patients would be more likely to receive surgery, and the intention was to avoid having surgery dominate the cost results. Despite this exclusion criterion, approximately 20–30% of patients in the study were judged by a central radiologist to have grade IV OA. It is not surprising that the grading provided by the site investigators and central radiologist differed for some patients, as the difference between grade level III and IV is subtle. The effectiveness of hylan G-F 20 is not expected to differ significantly for the two grade levels[27]. Patients with grade IV OA are also indicative of real world practice. To address the imbalance in X-ray grades between the two treatment groups, the analyses adjusted for Grade IV OA as a covariate; however, this did not change any results.

The demographics of aging populations make OA a particularly challenging medical and socioeconomic problem[28]. There is therefore a growing pool of patients with symptomatic knee OA who must be managed for many years, and in whom it is desirable to delay knee replacement for as long as possible. Currently the only treatments widely available for such patients are prescription NSAIDS or analgesics, intraarticular steroid injections, topical agents, and arthroscopic lavage and debridement. All of these available modalities have drawbacks or significant side effects. The data presented here indicate that the provision to patients with knee OA of viscosupplementation with hylan G-F 20 within an appropriate care treatment regimen provides benefits in the knee, overall health and health related quality of life at reduced levels of co-therapy and systemic adverse reactions.

Acknowledgments

The authors would like to thank individuals at Biomatrix, Inc. and Innovus Research Inc. who were involved in the study at various stages: Biomatrix, Inc.: Jean Bonomo, Katie Kubler, and Sarita Wilson. Innovus Research Inc.: Bill Bell, Lisa Bernard, Harminder Bilen, Linda Bursic, Helen Crabbe, Paule Desjardins, Julie Gordon, Susan Gow, Lorraine Hotchkies, Marg Hux, Natalie Milkovich, Nancy Ryan, Tammy Sarre, Tammy Slote, Sandra Therrien, John Walker, Irene Zivko, Barb Zrinscak. A special thanks to Shelley Lee for study coordination and Dan Pericak for statistical analysis.

References

1. MacLean CH, Knight K, Paulus H, Brook RH, Shekelle PG. Costs attributable to osteoarthritis. J Rheumatol 1998;25:2213–8.
2. Tugwell P. Economic evaluation of the management of pain in osteoarthritis. Drugs 1996;52(Suppl. 3):48–58.
3. Impact of Arthritis and Other Rheumatic Conditions on the Health-Care System-United States, 1997. Morbidity and Mortality Weekly Report 1997;48:349–53.
4. Felts W, Yelin E. The economic impact of the rheumatic disease in the United States. J Rheumatol 1989;16: 867–84.
5. Hochberg MC, Altman RD, Brandt KD, Clark BM, Dieppe PA, Griffin MR, *et al.* Guidelines for the medical management of osteoarthritis. Part II. Osteoarthritis of the knee. Arthritis Rheum 1995;38: 1541–6.
6. Singh G, Rosen Ramey DR. NSAID induced gastrointestinal complications: The ARAMIS perspective-1997. J Rheumatol 1998;Suppl 51 8–16.
7. Balazs EA, Denlinger JL. Viscosupplementation: a new concept in the treatment of osteoarthritis. J Rheumatol 1993;Suppl 39 3–9.
8. Berkowitz D. Proceedings of the United States Food and Drug Administration Advisory Panel on Orthopedic and Rehabilitation Devices 11/20/1996. Farfax, VA: CASET Associates 1996.
9. Towheed TE, Hochberg MC. A systematic review of randomized controlled trials of pharmacological therapy in osteoarthritis of the knee, with an emphasis on trial methodology. Semin Arthritis Rheum 1997;26:755–70.
10. Recommendations for the medical management of osteoarthritis of the hip and knee: 2000 update. American College of Rheumatology Subcommittee on Osteoarthritis Guidelines. Arthritis Rheum 2000; 43:1905–15.
11. Revicki DA, Frank L. Pharmacoeconomic evaluation in the real world. Effectiveness versus efficacy studies. Pharmacoeconomics 1999;15:423–34.

12. Bombardier C, Maetzel A. Pharmacoeconomic evaluation of new treatments: efficacy versus effectiveness studies? Ann Rheum Dis 1999;58:182–5.
13. Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic evaluations to the BMJ. The BMJ Economic Evaluation Working Party. BMJ 1996;313:275–83.
14. Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. J Chron Dis 1967;20: 637–48.
15. Canadian Coordinating Office for Health Technology Assessment. Guidelines for economic evaluation of pharmaceuticals: Canada—2nd ed. Ottawa: CCOHTA November 1997 (quote is from section 6.1, page 20).
16. Torrance GW, Raynauld JP, Walker V, Goldsmith CH, Bellamy N, Band PA, *et al*, In collaboration with the Canadian Knee OA Study Group. A prospective, randomized, pragmatic, health outcomes trial evaluating the incorporation of hylan G-F 20 into the treatment paradigm for patients with knee osteoarthritis (Part 2 of 2): economic results. Osteoarthritis Cart 2002; 10:518–27.
17. Kellgren JH, Lawrence JS. Osteoarthritis and disk degeneration in an urban population. Ann Rheum Dis 1958;17:388–97.
18. Bellamy N, Campbell J, Hill J, Goldman A, Band P. Validation of telephone-administration of the WOMAC Osteoarthritis Index. Abstract 1079. 14th European League Against Rheumatism Congress Abstract Book p 256.
19. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. A comparative study of telephone versus onsite completion of the WOMAC 3.0 Osteoarthritis Index. J Rheumatol 2002;29:783–6.
20. Goldsmith CH, Boers M, Bombardier C, Tugwell P. Criteria for clinically important changes in outcomes: development, scoring and evaluation of rheumatoid arthritis patient and trial profiles. OMERACT committee. J Rheumatol 1993;20:561–5.
21. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey. (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473–83.
22. Torrance GW, Furlong W, Feeny D, Boyle M. Multi-attribute preference functions. Health utilities index. Pharmacoeconomics 1995;7:503–20.
23. Little RJA, Rubin DB. Statistical Analysis with Missing Data. Toronto, ON: J Wiley & Sons Inc 1987.
24. Synvisc Hylan G-F 20 US label. CI 5014-1. 8005US-01. Issued August 28. 1997.
25. Felson DT. The epidemiology of knee osteoarthritis: results from the Framingham Osteoarthritis Study. Semin Arthritis Rheum 1990;20(3 Suppl 1):42–50.
26. March LM, Cross MJ, Lapsley H, Brnabic AJ, Tribe KL, Bachmeier CJ, *et al*. Outcomes after hip or knee replacement surgery for osteoarthritis. A prospective cohort study comparing patients' quality of life before and after surgery with age-related population norms. Med J Aust 1999;171:235–8.
27. Lussier A, Cividino AA, McFarlane CA, Olszynski WP, Potashner WJ, De Medicis R. Viscosupplementation with hylan for the treatment of osteoarthritis: findings from clinical practice in Canada. J Rheumatol 1996; 23:1579–85.
28. Helmick CG, Lawrence RC, Pollard RA, Lloyd E, Heyse SP for the National Arthritis Data Workgroup. Arthritis and other rheumatic conditions: who is affected now, who will be affected later? National Arthritis Data Workgroup. Arthritis Care Res 1995; 8:203–11.
29. Bellamy N, Campbell J, Hill J, Band P. A comparative study of of telephone versus onsite completion of the WOMAC 3.0 Osteoarthritis Index. J Rheumatol 2002; 29:783–6.

Osteoarthritis and Cartilage (2002) 10, 518–527

1063–4584/02/$35.00/0
doi:10.1053/joca.2001.0513, available online at http://www.idealibrary.com on IDEAL®

Osteoarthritis and Cartilage | CRS International Cartilage Repair Society | OARSI OSTEOARTHRITIS RESEARCH SOCIETY INTERNATIONAL

A prospective, randomized, pragmatic, health outcomes trial evaluating the incorporation of hylan G-F 20 into the treatment paradigm for patients with knee osteoarthritis (Part 2 of 2): economic results

G. W. Torrance*†, J. P. Raynauld‡, V. Walker†, C. H. Goldsmith*, N. Bellamy§, P. A. Band‖, M. Schultz¶ and P. Tugwell**, in collaboration with the Canadian Knee OA Study Group††
**Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada*
†Innovus Research Inc., Burlington, Ontario, Canada
‡Department of Medicine, University of Montreal, Montreal, Quebec, Canada
§Department of Medicine, The University of Queensland, Queensland, Australia
‖Biomatrix, Inc., Ridgefield, New Jersey, U.S.A.
¶Rhône-Poulenc Rorer Canada Inc., St Laurent, Quebec, Canada
***Department of Medicine, University of Ottawa, and Ottawa Hospital, Ottawa, Ontario, Canada*

Summary

Objective: Viscosupplementation with hylan G-F 20 has recently become registered for treatment of patients with osteoarthritis (OA) of the knee in most parts of the world. The cost effectiveness and cost utility of this new therapeutic modality were determined as part of a Canadian prospective, randomized, 1-year, open-label, multicentered trial.

Design: A total of 255 patients were randomized to 'Appropriate care *with* hylan G-F 20' (AC+H) or 'Appropriate care *without* hylan G-F 20' (AC). Costs (1999 Canadian dollars) were collected from the societal viewpoint and included all costs related to OA of the knee and OA in all joints. Patients completed a number of outcomes questionnaires including the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) and the Health Utilities Index Mark 3 (HUI3). Data were collected at clinic visits (baseline, 12 months) and by telephone (1, 2, 4, 6, 8, 10, and 12 months).

Results: The AC+H group over the year had higher costs ($2125–$1415=$710, $P<0.05$), more patients improved (69%–40%=29%, $P=0.0001$), greater increases in HUI3 (0.13–0.03=0.10, $P<0.0001$) and increased quality-adjusted life years (QALYs) (0.071, $P<0.05$). The incremental cost effectiveness ratio was $2505/patient improved. The incremental cost–utility ratio was $10 000/QALY gained. Sensitivity analyses and a second cost perspective gave similar results.

Conclusion: The cost–utility ratio is below the suggested Canadian adoption threshold. The results provide strong evidence for adoption of treatment with hylan G-F 20 in the patients and settings studied in the trial. © 2002 OsteoArthritis Research Society International. Published by Elsevier Science Ltd. All rights reserved.

Key words: Hylan G-F 20, Osteoarthritis, Knee, Economics, Cost and cost analysis, Health-related quality of life.

Introduction

Hylan G-F 20 (Synvisc® Genzyme Corporation, Cambridge, MA, U.S.A.) is a high-molecular weight viscosupplementation product for injection into the intraarticular space of the knee as a synovial fluid replacement. The product has molecular weight and viscosity similar to the synovial fluid found in healthy knees[1]. Hylan G-F 20 has been recently approved for the treatment of patients with osteoarthritis (OA) of the knee in most countries in the world. Accordingly,

Received 28 June 2001; revision requested 5 November 2001; revision received 22 November 2001; accepted 6 December 2001.
The study was funded by Biomatrix, Inc and Rhône-Poulenc Rorer Canada Inc.
Address correspondence to: Dr George Torrance, 1016-A Sutton Drive, Suite 200, Burlington, ON, Canada L7L 6B8. Tel: (905) 331 9911; Fax: (905) 331 9912; E-mail: torrance@mcmail.mcmaster.ca
††The Canadian Knee OA Study Group included the Steering Committee and clinical investigators. The Steering Committee consisted of J. P. Raynauld and G. W. Torrance, co-principal investigators; N. Bellamy, C. H. Goldsmith and P. Tugwell, co-investigators; P. A. Band, M. Schultz and V. Walker. The clinical investigators were: M. J. Bell, MD, Department of Medicine, University of Toronto, Toronto, ON; R. B. Bourne, MD, Department of Surgery, University of Western Ontario, London, ON; R. Cloutier, MD, Departement D'Orthopedie, Centre Hospitalier Affillie, Quebec, QC; A. M. Crawford, MB, Department of Medicine, University of Calgary, Calgary, AB; J. G. Cinats, Department of Surgery, University of Alberta Hospitals, Capital Health Regional Health Authority, Edmonton, AB; M.-A. Fitzcharles, MD, Department of Medicine, McGill University, Montreal, QC; R. D. Inman, MD, Division of Rheumatology, Toronto Western Hospital, University of Toronto, Toronto, ON; K. W. Marshall, MD, Division of Orthopaedics, Toronto Western Hospital, University of Toronto, Toronto, ON; D. P. O'Hanlon, MB, Department of Rheumatology, University of British Columbia, North Vancouver, BC; W. P. Olszynski, MD, Department of Medicine, University of Saskatchewan, Saskatoon, SK; J. Pope, MD, Department of Medicine, University of Western Ontario, London, ON; J.-P. Raynauld, MD, Department of Medicine, University of Montreal, QC; E. D. Sutton, MD, Department of Medicine, Dalhousie University, Halifax, NS; J. C. Thorne, MD, Department of Internal Medicine and Rheumatology, York County Hospital, Newmarket, ON.

the question we sought to answer in this research project was, given that this is an approved and used treatment, how effective and cost effective is it in the real world compared to appropriate care without its availability? That is, we sought to compare a world *with* hylan G-F 20 to a world *without* hylan G-F 20.

Health-related quality of life (HRQOL), effectiveness, cost-effectiveness and cost–utility are studied and reported here. The clinical and safety results of this study are reported in the accompanying manuscript[2]. The study was conducted following the Canadian guidelines for health economic studies[3], which in turn are consistent in most respects with similar guidelines in other countries[4–6]. This is a pragmatic trial. To enhance the real world generalizability of the results the study was conducted in multiple sites, with both rheumatologists and orthopedic surgeons, the study was 1 year in length, the inclusion/exclusion criteria were liberal, and the study was not blinded. The comparator was deliberately selected as appropriate care, not usual care. It was felt that usual care might contain some inappropriate care, and demonstrating that a new treatment is effective and cost-effective compared to inappropriate care is not particularly useful. Appropriate care is the preferred management strategy of specialists, rheumatologists or orthopedic surgeons, encouraged to follow the treatment guidelines published by the American College of Rheumatology[7], and instructed to treat conservatively.

The study was funded jointly by Biomatrix, Inc and Rhône-Poulenc Rorer Canada Inc. Innovus Research Inc., an independent contract research organization (CRO), was contracted to manage the study. An independent Steering Committee was assembled with the responsibility to design the study, develop the analysis plan, and disseminate study results. The Committee consisted of five academics, one representative from each of the two sponsoring companies and one representative from the CRO. The Steering Committee was deliberately structured to be dominated by the five independent academics on the Committee. The Steering Committee was very active and, in fact, dealt with all scientific questions that arose throughout the course of the study, and did so blinded to implications. The contractual arrangement gave the investigators unrestricted rights to publish the study results.

There are several audiences for the study. Clinicians will be interested in the findings of clinical effectiveness and of HRQOL. Many clinicians will also be interested in the findings of the cost-effectiveness and cost–utility analyses, particularly clinicians interested in the efficient use of limited resources and those involved in establishing treatment guidelines. Third-party payers, formulary managers, and fiscal administrators will be interested in all of the findings but particularly the results of the cost-effectiveness and cost–utility analyses.

Methods

STUDY DESIGN

This was a multicenter, 1-year, prospective, randomized, open-label, parallel design trial of appropriate care *with* hylan G-F 20 (AC+H) compared to appropriate care *without* hylan G-F 20 (AC) in the treatment of patients with symptomatic OA of the knee. Patients were recruited from 14 sites across Canada, 10 rheumatologists and four orthopedic surgeons. Patients had to be older than 40 years of age, to have a primary diagnosis of radiologically verified OA in the study knee (knee most symptomatic or with the most predominant musculoskeletal problem), excluding grade IV; to be symptomatic (total pain score greater than 175 mm on the five 100 mm visual analogue pain questions in the Western Ontario McMaster University Osteoarthritis Index (WOMAC)[8] despite prior treatment with acetaminophen or non-steroidal antiinflammatory drugs (NSAIDS) at any point prior to the study, and to be ambulatory.

Protocol-driven costs and outcomes were minimized by limiting study-induced clinic visits. Patients were assessed at the site during the baseline visit and the 12-month termination visit. Patients randomized to AC+H returned to the site for 2 consecutive weeks after baseline for the remaining hylan G-F 20 injections. Other visits could occur on an 'as needed' basis for clinical deterioration, treatment of adverse events, change in medication, or additional treatment with hylan G-F 20 if required; however, no other visits were required by the protocol.

Structured telephone interviews of the patients in both treatment groups were conducted by the CRO at 1, 2, 4, 6, 8, 10, and 12 months. The 12-month termination visit was included for patient assessment by the investigator and for measuring change since baseline. At the baseline visit the following data were collected: patient demographics, appropriate care treatment for knee OA, treatment for overall OA, concomitant medications, and patient self-administered questionnaires (WOMAC Likert 3.0)[9] 4-week recall, the Short-Form 36 (SF-36)[9] 4-week recall, and the Health Utilities Index 3 (HUI3)[10] 4-week recall. Except for patient demographics, the same information was collected at each telephone interview, with the addition of pill counts, adverse events, health care resources (e.g., physician visits, physiotherapy, hospitalizations), any patient expenses (e.g., travel), and lost time from work or usual activities due to OA treatment or OA symptoms. At each telephone interview the information was collected for the time period since the last interview, except for the patient self-administered questionnaires at months 4, 6, 8 and 12 where the recall period was 4 weeks. During the telephone interviews, the patient referred to the self-administered questionnaire and provided his/her answers to the telephone interviewer. To blind the patient to his/her previous answers to the same questions, s/he was instructed not to record the answers, and the questionnaire was laminated with plastic to make it difficult if someone tried to do so. Information collected during the telephone interviews (with the exception of the questionnaires) was compared with the patient's medical chart during monitoring visits and differences were resolved. The investigator reviewed the adverse events for possible attribution to study interventions.

OUTCOME MEASURES

The WOMAC Likert 3.0 is a disease-specific HRQOL instrument that asks the patient questions concerning the study knee. It produces an aggregate total score and scores for three subscales: pain, stiffness and physical functioning.

The outcome measure for the cost-effectiveness analysis (CEA) was patients improved. In the design of the study the Steering Committee provided two definitions of an improved patient. The primary definition was a patient whose WOMAC pain score at month 12 was reduced by

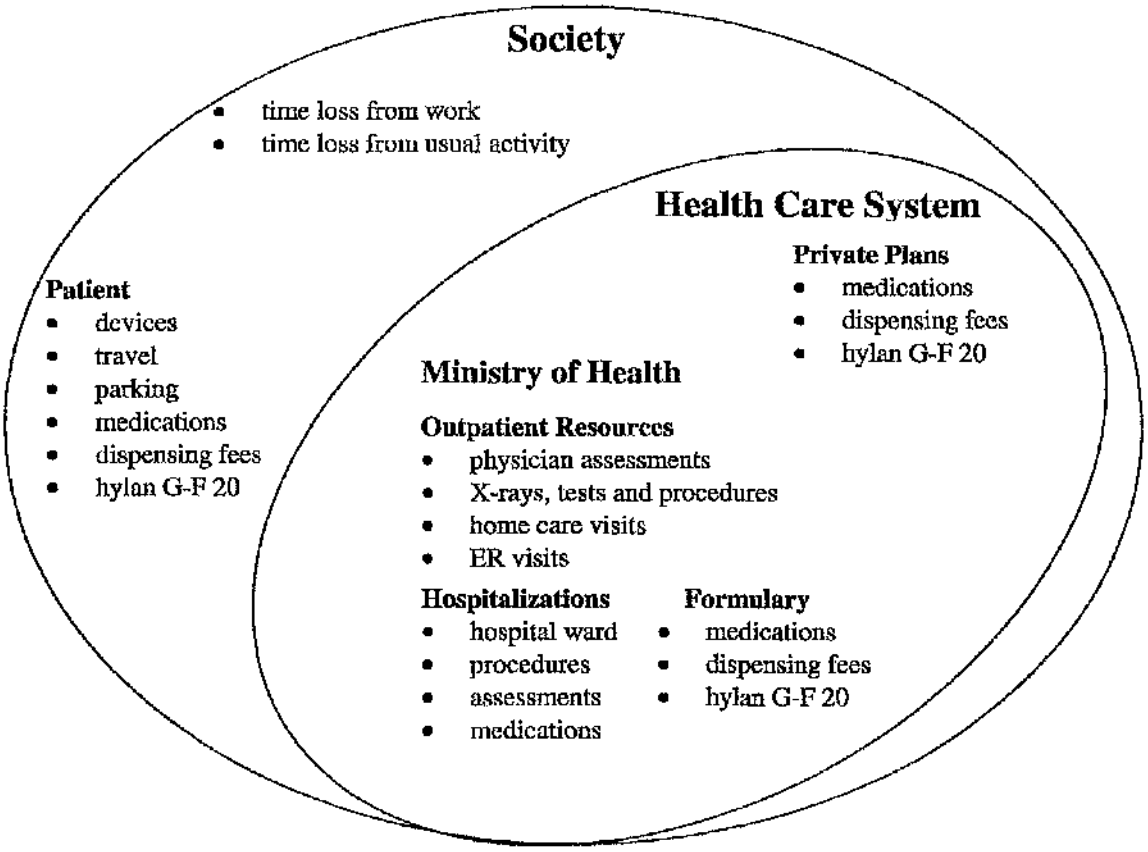


Fig. 1. The types of costs that were included in the societal and health care system perspectives.

20% or more compared with baseline. The secondary definition was a patient who not only reduced their pain score by 20% or greater but also reduced either their stiffness or their physical functioning score by 20% or more as well. The design also specified that the percentage of patients improved in the AC+H group would have to exceed the percentage in the AC group by at least 20% for the results to be clinically important.

The HUI3 is a generic, preference-weighted health status instrument that asks the patient questions about their overall health status and HRQOL. Specifically, the HUI3 measures health status using the following eight attributes: vision, hearing, speech, ambulation, dexterity, emotion, cognition and pain/discomfort. The patient is classified on each attribute into a level varying from normal to severely impaired. The scoring formula for the instrument is based on community preferences as measured by the standard gamble method and thus represents a von Neumann–Morgenstern utility[11]. The instrument provides an overall utility score (min: −0.36; max: 1) on the conventional health utility scale where dead=0.00 and perfect health=1.00. States worse than death can take on negative scores.

The outcome measure for the cost-utility analysis (CUA) is the number of quality-adjusted life years (QALY) gained. The overall utility score from the HUI3 is used as the quality adjustment factor for calculating QALYs gained. Note that the cost-effectiveness analysis focuses on the *study knee* effectiveness, whereas the cost-utility analysis focuses on *patient* effectiveness.

PERSPECTIVES

Figure 1 shows the categories of costs that are included in the different perspectives. Some costs, such as medications and hylan G-F 20, fall in more than one perspective, depending upon the patient's drug plan. A comprehensive societal perspective was adopted as the primary perspective for the economic analyses. In this perspective all costs are counted. Lost time was captured for both the patient and for the unpaid family caregiver, and was categorized into lost work time (for those in paid employment) and lost usual activity time (for those not in paid employment). In the base case analysis only lost work time was included. In a sensitivity analysis, all lost time was included. The health care system (HCS) consists of the two major payers in Ontario, Ministry of Health and private medical plans. This perspective was adopted as the secondary perspective.

RESOURCE UTILIZATION AND COSTING

At each telephone interview patients reported health care received, and indicated which they thought were related to OA (i.e., due to OA in any joint, the treatment of OA in any joint, or the treatment of adverse events related to the treatment of OA). The patient's data were compared with the patient's chart at the investigator's office, and discrepancies were resolved. The physician or the research coordinator at the site reviewed the items and could override the patient's attribution to OA. To improve

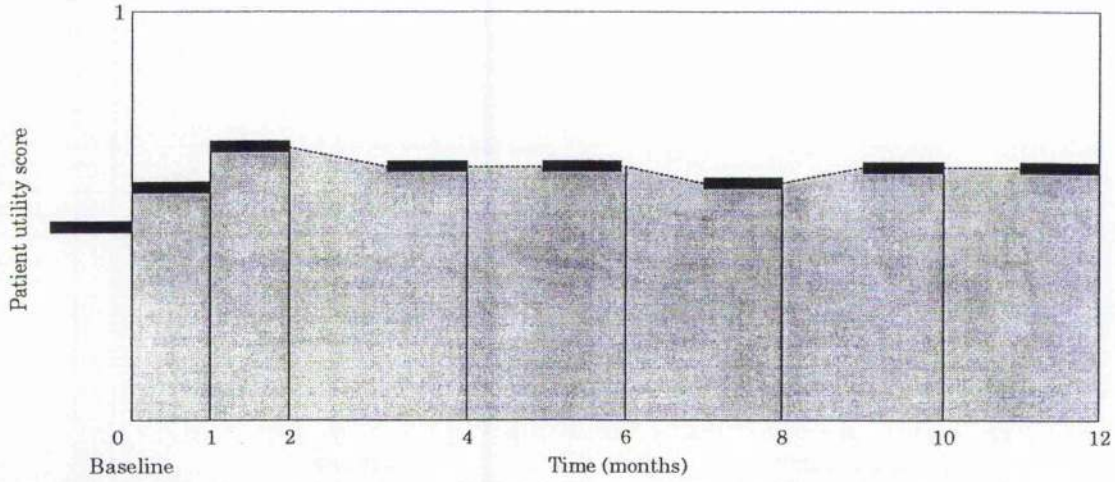


Fig. 2. A typical patient profile for the HUI3 score showing the area-under-the-curve calculation for quality-adjusted life years (QALY). The baseline HUI3 score is shown on the y-axis; however, it is not used to calculate the QALY.

consistency across sites, the three clinicians (JPR, PT, NB) from the Steering Committee reviewed the attribution of all resources, while blinded to treatment allocation. They could override the site's attribution to OA. Only costs related to OA were included in the analysis. The following protocol driven items were not included in the costs: screening visit, X-ray at screening, lab test at baseline, and termination visit.

Costs are reported in 1999 Canadian dollars ($Can). Costs not available in 1999 dollars were adjusted to 1999 using the health and personal care component of the consumer price index[12]. Costs are from the province of Ontario, Canada's largest province.

The market price of hylan G-F 20 in Canada during the study was $339 including tax per course of three injections. Medications were priced according to the Best Available Price for drugs listed on the Ontario Drug Benefit Formulary, or the brand name price from a pharmacy wholesaler's catalogue[13,14]. Prices for outpatient resources were obtained from a variety of appropriate sources; e.g., fees for physician services and laboratory and procedures[15], and cost of other health care professionals[16]. A standard cost ($165.55) for a generic emergency room (ER) visit was used for all ER visits in the study[16]. The reason for hospitalization (e.g. total knee replacement) was coded into an international classification of disease ninth revision clinical modification (ICD9-CM) code. The mean cost for patients hospitalized for that ICD9-CM code for the same length of stay was employed[17]. Patient lost productivity was valued at the Canadian average industrial wage rate ($121.59 per day) for time lost from employment and non-work time losses[18].

STATISTICS

The sample size for the study was calculated based on the primary effectiveness measure for the clinical results, mean change in WOMAC pain score in study knee, as described in the accompanying paper[2]. The sample size was not calculated on the basis of the cost-effectiveness or cost–utility ratio because well-established methods to do so did not exist at the time the study was designed.

All patients randomized were included in the analysis. Missing data were imputed, so that a full data set was available for the statistical analyses. The hot deck method was used to impute data[19]. A patient with missing data was matched to a small group of 'similar' patients with complete data, from which one was selected randomly. Data points missing in the index patient were filled in from the matched patient. Consistent with the comparison of a world with hylan G-F 20 to a world without hylan G-F 20, the few patients in the AC group who violated the protocol by receiving hylan G-F 20 treatment were treated as drop-outs at that point[2]. That is, their data from that point forward were imputed, just like any other drop-out.

The WOMAC and HUI3 questionnaires specified a recall period of 4 weeks, except for the questionnaires at months 1 and 2 in which the recall period was the time since the previous visit. The HUI3 overall utility score represents the mean score for that patient over the recall time period. A typical patient profile is shown in Fig. 2. Note, the horizontal segments in Fig. 2 represent measured scores while the sloping segments represent linear interpolation. The QALY for each patient is calculated by taking the area under the curve for the patient's utility, using years as the unit for time.

Because all patients were not in the study for exactly 365 days, their costs and QALYs were converted to an equivalent annual figure [annualized cost or QALY=(total cost or QALY for time in study/number of days on study)×365.25 days]. Because the time horizon for the analysis was 1 year, discounting of future costs and consequences was not necessary.

The base case analysis is the primary analysis. The following one-way sensitivity analyses were performed to test the robustness of the results:

- Effectiveness: for the CEA, the incremental effectiveness (difference in proportion of patients improved) was varied to its upper and lower 90% confidence bounds. Similarly, for the CUA, the incremental effectiveness (QALYs gained) was varied to its upper and lower 90% confidence bounds.
- Cost: for the CEA and CUA, the incremental cost was varied to its upper and lower 90% confidence bounds.

Table I
*Demographic information and osteoarthritis status, f (percent of n)**

	AC+H (*n*=127)	AC (*n*=128)
Age, mean (s.d.) years	62.6 (9.4)	63.5 (10.5)
Sex, female	86 (68%)	93 (73%)
Work status		
Full-time	30 (24%)	19 (15%)
Part-time	11 (9%)	16 (13%)
Sick leave	1 (1%)	2 (2%)
Not in paid employment	84 (66%)	90 (70%)
Not specified	1 (1%)	1 (1%)
Prescription drug plan coverage		
No plan	15 (12%)	15 (12%)
Employer or private	53 (41%)	39 (31%)
Government	47 (37%)	64 (50%)
Government+(private or employer)	11 (9%)	8 (6%)
Not specified	1 (1%)	2 (2%)
Duration (years) of OA symptoms		
Study knee, mean (s.d.)	9.0 (9.5)	9.9 (9.7)
Other knee, mean (s.d.)	7.4 (8.8)	8.3 (9.3)
OA at baseline		
Other knee affected	109 (86%)	108 (84%)
Other joints affected	86 (68%)	78 (61%)

*f is frequency, *n* is sample size. Not all percentages sum to 100 due to rounding.
OA=osteoarthritis; AC+H=Appropriate Care + hylan G-F 20; AC=Appropriate Care.

- Time loss: the CEA and CUA were re-done using a more liberal costing of time loss that costed all time loss whether work or usual major activity (including leisure). In the base case only time loss from work was costed.

Results

PATIENTS

One hundred and twenty-seven patients were randomized to receive AC+H and 128 to receive AC. The demographic and OA status of the patients are displayed in Table I. The patients had a mean age of 63 years, with the preponderance of them being unemployed women with OA in both knees, and covered by a drug plan. The two groups were well balanced.

COSTS

The mean annual OA-related cost per patient from the societal perspective by type of cost is shown in Table II. There were too many different kinds of costs to show unit costs within type (e.g. 20 types of injections, 80 types of outpatient resources, 30 types of assistive devices). The total annual cost per patient in a world without hylan G-F 20 (AC group) was $1415. The total in a world with hylan G-F 20 was $2125, an excess of $710. The 95% confidence

Table II
Mean annual OA-related cost per patient from the societal perspective

	AC+H (*n*=127)		AC (*n*=128)		Mean difference (AC+H)−AC)
hylan G-F 20	676.01	(370.89)	0.00	(0.00)	676.01
Knee OA appropriate care treatment					
Injections (e.g. corticosteriods)	4.05	(10.42)	18.45	(17.31)	−14.40
Medications (e.g. NSAIDs)	200.63	(242.95)	370.10	(529.13)	−169.48
Other therapy (e.g. physiotherapy)	237.32	(831.22)	305.10	(669.22)	−67.78
Assistive devices (e.g. cane)	5.38	(11.61)	16.38	(54.58)	−11.00
Procedures (arthroscopy)	1.70	(19.19)	18.12	(118.65)	−16.42
Subtotal (knee treatment)	449.08		728.15		−279.08
Concomitant medications					
OA in other joints (e.g. NSAIDs)	16.91	(67.62)	17.81	(72.06)	−0.90
Adverse events due to OA treatment (e.g. antacid, analgesics)	53.88	(179.05)	50.08	(123.68)	3.80
Subtotal (concomitant meds)	70.79		67.89		2.90
Outpatient resources (e.g. physician visits)	245.72	(399.96)	134.02	(135.11)	111.70
Hospitalization	194.53	(1012.28)	101.57	(752.54)	92.96
Time loss from work					
by patient					
Due to OA	229.13	(942.26)	190.37	(1085.09)	38.76
Due to OA treatment	53.49	(150.86)	37.95	(180.66)	15.54
by caregiver					
Due to OA	0.37	(4.15)	0.06	(0.69)	0.31
Due to OA treatment	35.30	(350.30)	6.56	(26.32)	28.74
Subtotal (time loss)	318.29		234.94		83.35
Out-of-pocket expenses (e.g. transportation)	170.30	(250.09)	148.00	(215.23)	22.30
Total cost	2124.71	(2528.35)	1414.58	(2032.74)	710.13

All costs are in 1999 Canadian dollars. Mean (s.d.).
OA=osteoarthritis; AC+H=Appropriate Care+hylan G-F 20; AC=Appropriate Care; NSAIDs=non-steroidal antiinflammatory drugs.

interval for the difference in mean total costs was a lower bound of $147 and an upper bound of $1273. Thus, the cost difference between groups was statistically significant at the 5% level (95% confidence interval did not include 0). From a HCS perspective, the total annual OA-related cost was also greater in the AC+H group, and by almost the same amount. The difference was $705 which was also statistically significant at the 5% level (data not shown).

The major contributor to the societal incremental cost of $710 was the cost of the hylan G-F 20 itself, $676. This was the average cost of hylan G-F 20 per patient over the year in the AC+H group. The actual cost of the product for a treatment of three injections was $339, but because many patients had the other knee done as well, and some had additional treatments throughout the year, the average cost was $676. The second major contributor to the incremental cost of $710 was a savings in other treatment costs for the knee OA of $279 (Table II, knee OA appropriate care treatment). The third major contributor was the $112 extra for outpatient visits, primarily the visits to receive injections of hylan G-F 20. At $93 hospitalizations were the next largest contributor. In the base case analysis there were a total of five hospitalizations attributable to OA in the AC+H group and three in the AC group. The five in the AC+H group were: total knee replacement in study knee, total knee replacement in other knee, total hip replacement, triple ankle fusion, and tibia osteotomy. The three in the AC group were: total knee replacement in study knee, total knee replacement in other knee, and bunionectomy. Interestingly, there were two additional total knee replacements in the study knee that were not counted in the base case analysis because they occurred after the two patients in question had violated protocol by receiving hylan G-F 20. The fifth largest contributor to the cost difference was the additional cost of lost work time for the AC+H group at $83, which could be due to the visits needed for the hylan G-F 20 injections. Out of pocket expenses were also slightly higher possibly for the same reason, i.e., travel costs for visits.

CONSEQUENCES

The percent of patients improved at 12 months using the primary definition of improvement was 69% in the AC+H group and 40% in the AC group for an increment of 29%. Using the secondary definition of improvement the results were 62% and 35% for an increment of 27%. Both increments were statistically significant (P=0.0001) and exceeded the clinically important difference of 20% established a priori as part of the research design.

The improvement in mean utility from baseline to termination as measured by the HUI3 was 0.13 in the AC+H group compared to 0.03 in the AC group, for a difference of 0.10 units of utility (P<0.0001). Figure 3 displays the change in mean utility score from baseline to each interview for both treatment groups. Both groups improved sharply for the first 2 months and then tailed off. However, the AC+H group improved more and tailed off less, thus giving a substantial area between the two curves. The area between the two curves over the 12 months represents the difference in QALYs between the two groups. The patients in the AC+H group gained 0.071 QALYs compared to the patients in the AC group. The 95% confidence interval for the difference in QALYs was a lower bound of 0.017 and an upper bound of 0.126. The difference between groups was statistically significant at the 5% level (95% confidence interval did not include 0).

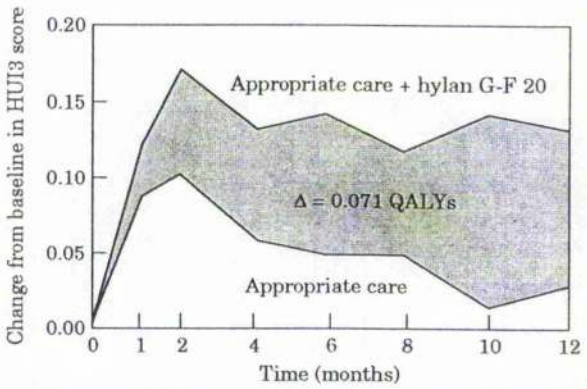


Fig. 3. Difference between treatment groups in mean change from baseline in Health Utilities Index 3 utility score. HUI3=Health Utilities Index 3; QALYs=quality-adjusted life years.

BASE CASE ANALYSIS

The base case CEA and CUA are shown in the first row of Table III. The AC+H group was more costly and more effective. The incremental cost per patient over 1 year was $710 and $705 from the societal and HCS perspective, respectively. The incremental effectiveness was an increase of 0.2834 proportion of patients improved. The C/E ratio was $2505 or $2488 per patient improved, from the societal and HCS perspective, respectively. For the CUA the incremental effectiveness was 0.071 QALYs per patient. The C/U ratio was $10 000 or $9930 per QALY gained, from the societal and HCS perspective, respectively.

SENSITIVITY ANALYSES

Table III displays the results of the sensitivity analyses for the CEA and CUA. There are five sensitivity analyses: effectiveness high, effectiveness low, costs high, costs low, and alternative definition of lost time. To help interpret these results the CUA sensitivity analyses are plotted on a cost-effectiveness graph in Fig. 4. Note that the slope of the line through the point is the cost per QALY of that point. Thus, lower slopes are more cost-effective, and vice versa. To enhance the interpretation we have also plotted on Fig. 4 the decision thresholds suggested by Laupacis *et al.*[20]: cost per QALY between $0 and $20 000=strong evidence for adoption; between $20 000 and $100 000=moderate evidence for adoption; and above $100 000=weak evidence for adoption. Figure 4 demonstrates that the results are robust; four of the five sensitivity analyses fall in the decision sector 'strong evidence for adoption' while the fifth falls in the adjacent sector 'moderate evidence for adoption'.

Discussion

Although the trial was powered only for the primary clinical outcome (change in mean WOMAC pain score), the outcomes for the economic evaluation (gain in percent of patients improved, gain in QALYs, and cost difference) also achieved statistical significance at the 5% level. Thus the findings are particularly robust, especially for a prospective economic evaluation study.

Table III
Cost-effectiveness and cost- utility analyses: base case and sensitivity analyses from the societal and health care system perspectives

	Annual cost difference		Difference in proportion of patients improved	Cost per patient improved*		QALYs gained†	Cost per QALY gained‡	
	Societal	HCS		Societal	HCS		Societal	HCS
Number of patients (AC+H/AC)§	127/128	127/128	127/127	127/128	127/128	127/128	127/128	127/128
Base-case analysis	$710	$705	0.2834	$2505	$2488	0.071	$10 000	$9930
Sensitivity analyses on outcomes‖								
High	$710	$705	0.3820	$1859	$1846	0.117	$6068	$6026
Low	$710	$705	0.1848	$3842	$3815	0.025	$28,400	$28,200
Sensitivity analyses on costs‖								
High	$1183	$1008	0.2834	$4174	$3557	0.071	$16,662	$14,197
Low	$238	$402	0.2834	$840	$1418	0.071	$3352	$5662
Sensitivity analysis using alternative costing for time loss**	$938	n/a††	0.2834	$3310	n/a††	0.071	$13,211	n/a††

All costs are in 1999 Canadian dollars.
*The cost per patient improved=incremental cost ([AC+H]−AC)/incremental effectiveness ([AC+H]−AC).
†QALY gained is adjusted for baseline differences.
‡The cost per QALY gained=incremental cost ([AC+H]−AC)/incremental QALY ([AC+H]−AC).
§The mean cost per patient and QALYs gained were calculated from 128 patients; the proportion of patients improved was calculated from 127 patients.
‖High represents the upper 90% CI for the difference between groups, and Low represents the lower 90% CI for the difference between groups.
**The base-case analysis included time loss from work. The sensitivity analysis included time loss from usual major activities also.
††A ratio will not be calculated as the cost of time loss was not included in the HCS perspective costs.
QALY=quality-adjusted life year; HCS=health care system; AC+H=Appropriate Care+hylan G-F 20; AC=Appropriate Care; CI=confidence interval.

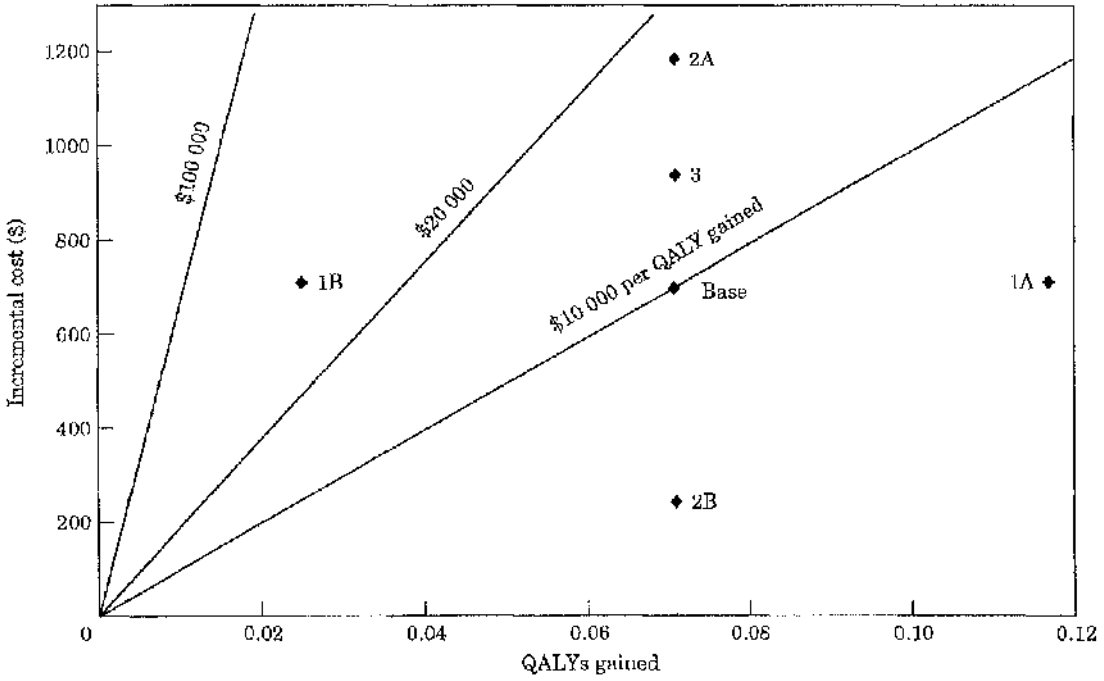


Fig. 4. Base-case (Base) and sensitivity analysis (1A-3) results for the cost–utility analysis from the societal perspective. Sensitivity analyses assume: 1A, upper 90% CI for difference in QALYs gained between groups; 1B, lower 90% CI for difference in QALYs gained between groups; 2A, upper 90% CI for difference in total costs between groups; 2B, lower 90% CI for difference in total cost between groups; 3, annual cost of time loss from work or usual major activity for all patients and caregivers. CI=confidence interval; QALY=quality-adjusted life year; $Can=Canadian dollars.

The cost findings are relatively easy to interpret. The use of hylan G-F 20 reduced the need for and the cost of other treatments for OA, but not enough to offset the increased costs due to the price of the product and due to the costs associated with the extra physician visits required to administer the treatment.

The patient outcome findings are also straightforward. The patients in the AC+H group were better off, and

statistically significantly so, by both the disease specific measure (WOMAC), and by the health utility measure (HUI3). The important outcomes for the economic analysis were the incremental proportion of patients improved in the AC+H group compared to the AC group (29%) and the incremental QALYs gained in the AC+H group compared to the AC group (0.071). The former outcome, 29%, exceeded the a priori threshold for clinical importance which had been set at 20%. However, no threshold had been established a priori for an important increment in QALYs.

Is a mean gain of 0.071 QALYs per patient important? One way to address this question, is to return to the theory on which the HUI3 instrument is based, von Neumann–Morgenstern utility theory and the standard gamble measurement. On the basis of this theory a direct interpretation is that a gain of 0.071 QALYs over a year is equivalent to a reduction in mortality rate of 0.071. That is, providing an *ongoing* improvement in quality of life of this magnitude (0.071 QALY per year) is equivalent to finding a group of healthy individuals (no HUI3 disabilities) who are at high risk (50%) of immediate sudden death and providing them with an absolute risk reduction in mortality of 7.1% (i.e., reducing their risk to 42.9%). If the group of individuals has compromised quality of life like patients with knee OA, the interpretation is even more dramatic. According to von Neumann–Morgenstern utility theory a patient like those at baseline in the AC+H group with a utility score of 0.50 and no risk of immediate death would be willing to take a risk of immediate death of 12.4%, 0.124=1.000−(0.500/(0.500+0.071)), to achieve an ongoing improvement in quality of life of 0.071. Thus, there is little question that a QALY gain of this magnitude is important. Moreover, any QALY gain can be important depending on the cost required to produce the gain and on the overall context of the gain[21].

The cost-effectiveness finding is that the incremental cost per patient improved over 1 year is $2505 (societal) and $2488 (HCS). That is, an expenditure of approximately $2500 will purchase an improved patient for a year. Is this good value for money? There is no absolute answer. Decision-makers responsible for allocating resources will have to weigh this opportunity against other choices. If the other choices are not expressed in the same metric, improved patients, the comparison becomes difficult. This highlights the advantage of cost–utility analysis.

The finding of the cost–utility analysis is that the incremental cost per QALY gained is $10 000 (societal) and $9930 (HCS). That is, an expenditure of approximately $10 000 will purchase a gain of 1 QALY. Is this good value for money? There are several ways to approach the question. One approach is to compare the results to other studies using a league table in which studies are ranked from best to worst according to their cost per QALY gained. Current methodological advice is that indiscriminate comparisons of this type can be misleading, and that league tables should be restricted to high-quality studies that use comparable scientific methods, and possibly further restricted to interventions targeted at one condition (e.g., musculo-skeletal problems)[22].

Chapman *et al.* from Harvard University undertook a comprehensive literature review of cost–utility studies 1976–1997 and categorized them into 'Panel-worthy' or not, and further subdivided the list by disease categories one of which is musculo-skeletal[23]. 'Panel-worthy' studies are those that met a minimum standard of methodological quality established by Chapman *et al.* based on the recommendations of the Panel on Cost-Effectiveness in Health and Medicine[24]. There has been only one 'Panel-worthy' study in the field of musculo-skeletal diseases, a cost-effectiveness analysis of total hip arthroplasty for OA of the hip, published in 1996 by Chang *et al.* In addition, there is a relevant 'Panel-worthy' study in the digestive system category, a cost-utility analysis of the use of misoprostol prophylaxis for rheumatoid arthritis patients receiving NSAIDs drugs, published by Gabriel *et al.* in 1994. These two studies contain four cost–utility ratios. The four ratios are included in our league table (Table IV).

Table IV
Cost/QALY League Table (1998 US dollars)

Cost/QALY gained	Treatment and comparator
Cost-saving	Total hip arthroplasty vs no total hip arthroplasty in white 60-year-old women with hip osteoarthritis in American College of Rheumatology function class III (significant functional limitation, but not dependent)[23]
$5500	Total hip arthroplasty vs no total hip arthroplasty in white men ≥85 years old with hip osteoarthritis in American College of Rheumatology function class III (significant functional limitation, but not dependent)[23]
$6500	Appropriate Care+hylan G-F 20 vs Appropriate Care for knee osteoarthritis, health care system perspective (this study)
$6600	Appropriate Care+hylan G-F 20 vs Appropriate Care for knee osteoarthritis, societal perspective (this study)
$7500	Total hip arthroplasty vs no hip arthroplasty for all patients, 3 year follow-up[25]
$11 000	Prophylaxis for NSAID-associated gastric ulcers with low-dose misoprostol (100 mcg four times daily) for elderly (>60 years old) vs no prophylaxis for all NSAID users in rheumatoid arthritis patients on NSAIDs[23]
$12 000	Prophylaxis for NSAID-associated gastric ulcers with low-dose misoprostol (100 mcg four times daily) for all vs prophylaxis for elderly (>60 years old) in rheumatoid arthritis patients on NSAIDs[23]

In addition we have included, for comparison, one relevant Canadian study[25]. Although it did not meet the criteria for 'Panel-worthy', we believe it can be usefully interpreted. One shortcoming of the study, lack of discounting, we have corrected in the data shown here using a discount rate of 5% per year. Another shortcoming, lack of incremental costing (they did not measure the costs that would have occurred without hip replacement) was conservative. On the positive side, the study was Canadian and, thus, is more directly comparable to our hylan G-F 20 study. The study prospectively measured costs (from the perspective of the health care system) and time trade-off utilities for total hip arthroplasty over 1 year, and modeled the analysis for 2 additional years for a total analytic horizon of 3 years.

For consistency with the Harvard table of 'Panel-worthy' studies, all entries in the league table (Table IV) have been adjusted to 1998 US dollars. The two Canadian ratios were first adjusted to 1998 Canadian dollars, using the Health Care component of the Canadian Consumer Price Index[26], and then converted to US dollars using the mean exchange rate for 1998, 1.4831[27]. As shown in the league table, hylan G-F 20 provides 'value for money', from either perspective, that is not as good as total hip arthroplasty for 60-year-old

US women as studied by Chang *et al.*, but is similar to total hip arthroplasty for ≥85-year-old US men studied by Chang *et al.* or for Canadians studied by Laupacis *et al.*[25], and is better than misoprostol prophylaxis for rheumatoid arthritis patients taking NSAIDs.

A second approach to answering the question of value for money is to compare the results to some external standard. For example, Laupacis *et al.*[20] suggest that if a new therapy is more effective and more costly than the existing one and costs less than $20 000/QALY gained, there exists strong evidence for adoption of the new therapy. Similarly, $20 000 to $100 000/QALY gained provides moderate evidence for adoption, and over $100 000/QALY gained provides weak evidence for adoption. These costs are in 1990 Canadian dollars, and the thresholds in 1999 dollars may be larger. Thus, the cost–utility results for the use of hylan G-F 20 in the knee fall in the category of strong evidence in favour of adoption.

STUDY STRENGTHS

The study was designed according to rigorous standards put forth in guidelines for economic evaluations. All of the participating investigators had a high degree of experience treating patients with hylan G-F 20, therefore it was possible to measure effectiveness, or 'real world' clinical effects, the outcome required for an economic evaluation. Hylan G-F 20 was compared to an appropriate comparator in that current practice guidelines were employed rather than placebo. The study was open-label and therefore physicians could practice according to their normal routine. Both treatment groups had equal access to the full repertoire of appropriate care. It was possible to limit the amount of protocol-driven costs by employing telephone interviews to collect data. Costs included indirect and direct costs, and overhead costs were included. The study time horizon was 1 year, which enabled measurement of downstream costs and consequences associated with subsequent courses of hylan G-F 20, adverse events, and treatment failure.

STUDY LIMITATIONS

One of the limitations was that the study was open-label and patients or physicians may have been biased in favour of hylan G-F 20 treatment which in turn could have affected outcomes and costs. On the other hand, many in the AC group also received a knee injection (corticosteroid), and all patients in both groups received appropriate care. The ongoing telephone interviews may have influenced the patient assessments. The patients might have improved because of the attention they received, or they might have felt worse because they focused more on their symptoms when answering symptom questionnaires. However, the potential telephone interview biases would have been similar in both groups, therefore it is unlikely there was a differential from this source.

GENERALIZABILITY

The study applies to patients treated by rheumatologists and orthopedic surgeons in Canada. The results are not directly applicable in other countries. Readers in other countries have to decide which aspects of the study apply, and which aspects need to be modified. It is generally felt that clinical findings travel fairly well. One would expect similar findings in the patient outcomes—WOMAC scores, SF-36 scores and HUI3 scores. Moreover, although the HUI3 is scored based on preferences from a Canadian population, there is considerable evidence that preference scores from the general public are independent of country (for example, see Johnson *et al.*[28] and Gales *et al.*[29]). On the other hand, because the health care systems differ among countries, the utilization of health care resources may differ. Moreover, the prices of health care resources including hylan G-F 20 differ in different countries. Thus, the costs can not be assumed to apply in other countries. Those interested in other countries will have to modify the study results appropriately[30] or use this study as a prototype from which to conduct a study in their own country.

Conclusions

This study demonstrates that hylan G-F 20, when used in conjunction with appropriate care, provides an improvement in outcomes that is both clinically important and statistically significant. Total costs are higher when hylan G-F 20 is selected as a treatment option for patients with OA of the knee, but the cost per QALY gained is well below the suggested Canadian threshold for adoption.

Acknowledgments

The authors would like to thank individuals at Biomatrix, Inc. and Innovus Research Inc. who were involved in the study at various stages: Biomatrix, Inc.: Jean Bonomo, Katie Kubler, and Sarita Wilson. Innovus Research Inc.: Bill Bell, Lisa Bernard, Harminder Bilen, Linda Bursic, Helen Crabbe, Paule Desjardins, Julie Gordon, Susan Gow, Lorraine Hotchkies, Marg Hux, Natalie Milkovich, Nancy Ryan, Tammy Sarre, Tammy Slote, Sandra Therrien, John Walker, Irene Zivko, Barb Zrinscak. A special thanks to Shelley Lee for study coordination and Dan Pericak for statistical analysis.

References

1. Synvisc Hylan G-F 20 US label. CI 5014-1. 8005US-01. Issued August 28, 1997.
2. Raynauld JP, Torrance GW, Band PA, Goldsmith CH, Tugwell P, Walker V, *et al.* in collaboration with the Canadian Knee OA Study Group. A prospective, randomized, pragmatic, health outcomes trial evaluating the incorporation of hylan G-F 20 into the treatment paradigm for patients with knee osteoarthritis (Part 1 of 2): clinical results. Osteoarthritis Cart 2002;10:506–17.
3. Canadian Coordinating Office for Health Technology Assessment. Guidelines for Economic Evaluation of Pharmaceuticals: Canada. 2nd ed. Ottawa: CCOHTA, November 1997.
4. Australian Commonwealth Department of Human Services and Health. Guidelines for the Pharmaceutical Industry on Preparation of Submissions to the Pharmaceutical Benefits Advisory Committee: including major submissions involving economic analyses. Canberra: Australian Government Publishing Service 1995.

5. Weinstein MC, Siegel JE, Russell LB, Gold MR. Recommendations for reporting cost-effectiveness analyses. Panel on cost-effectiveness in health and medicine. JAMA 1996;276:1339–41.
6. Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic evaluations to the BMJ. The BMJ Economic Evaluation Working Party. BMJ 1996;313:275–83.
7. Hochberg MC, Altman RD, Brandt KD, Clark BM, Dieppe PA, Griffin MR, *et al.* Guidelines for the medical management of osteoarthritis. Part II. Osteoarthritis of the knee. Arthritis Rheum 1995;38: 1541–6.
8. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988;15:1833–40.
9. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey. (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473–83.
10. Torrance GW, Furlong W, Feeny D, Boyle M. Multi-attribute preference functions. Health utilities index. Pharmacoeconomics 1995;7:503–20.
11. Von Neumann J, Morgenstern O. Theory of Games and Economic Behaviour. Princeton, NJ: Princeton University Press 1944.
12. Statistics Canada. (1995) Consumer Price Index, January 1996 to 1998 (http://www.statcan.ca/Daily). Ottawa, Ontario: Ministry of Industry.
13. Ontario Drug Benefit Formulary/Comparative Drug Index Edition No. 35.
14. Medis June 1998 Pharmaceuticals catalogue.
15. Ontario Schedule of Benefits. Physician Services under the Health Insurance Act. April 1, 1998.
16. Browne G, Gafni A, Roberts J, Goldsmith A, Jamieson E. Approach to the Measurement of Costs (Expenditures) when Evaluating Health and Social Programmes, System-Linked Research Unit, McMaster University. November, 1995.
17. Ontario Case Costing Project. 1994/1995 & 1995/96. 20 Adelaide St. East, Toronto, ON, M5G 2T6.
18. Statistics Canada. Average industrial aggregate. Ottawa (ON): Statistics Canada. February 1999.
19. Little RJA, Rubin DB. Statistical Analysis with Missing Data. Toronto, ON: Wiley 1987.
20. Laupacis A, Feeny D, Detsky AS, Tugwell PX. How attractive does a new technology have to be to warrant adoption and utilization? Tentative guidelines for using clinical and economic evaluations. CMAJ 1992;146:473–81.
21. Hays RD, Wooley JM. The concept of clinically meaningful difference in health-related quality of life research: how meaningful is it? Pharmacoeconomics 2000;18:419–23.
22. Drummond M, Torrance G, Mason J. Cost-effectiveness league tables: more harm than good? Soc Sci Med 1993;37:33–40.
23. Chapman RH, Stone PW, Sandberg EA, Bell C, Neumann PJ. A comprehensive league table of cost-utility ratios and a sub-table of 'panel worthy' studies. Med Dec Mkg 2000;20:451–67.
24. Gold MR, Siegel JE, Russell LB, Weinstein MC, Eds. Cost-effectiveness *in* Health and Medicine. New York: Oxford University Press 1996.
25. Laupacis A, Bourne R, Rorabeck C, Feeny D, Wong C, Tugwell P, *et al.* Costs of elective total hip arthroplasty during the first year. Cemented versus non-cemented. J Arthroplasty 1994;9(5): 481–7.
26. Statistics Canada, CANSIM, Matrix 9940, Consumer price index, 1996 classification, annual health care, historical summary.
27. The Canada Customs and Revenue Agency. Average Exchange Rates for 1998.
28. Johnson JA, Ohinmaa A, Murti B, Sintonen H, Coons SJ. Comparison of Finnish and US-based visual analog scale valuations of the EQ-5D measure. Med Dec Mkg 2000;20:281–9.
29. Le Gales C, Buron C, Costet N, Rosman S, Slama G. The French Health Utilities Index Mark 3. Abstract. Value in Health 2000;3:103.
30. O'Brien B. A tale of two (or more) cities: geographic transferability of pharmacoeconomic data. Am J Managed Care 1997;3:S33–9.

Appendix

AC+H=Appropriate care *with* hylan G-F 20
AC=Appropriate care *without* hylan G-F 20
WOMAC=Western Ontario and McMaster Universities Osteoarthritis Index
HUI3=Health Utilities Index 3
QALYs=quality-adjusted life years
OA=osteoarthritis
HRQOL=Health-related quality of life
CRO=contract research organization
NSAIDs=non-steroidal antiinflammatory drugs
SF-36=Short-Form 36
HCS=health care system
$Can=Canadian dollars
ER=emergency room
ICD9-CM=international classification of disease ninth revision clinical modification
CEA=cost-effectiveness analysis
CUA=cost-utility analysis

SECTION 6 - Globalisation

The early phase of globalisation of the WOMAC Index occurred contemporaneous with two parallel but largely unrelated processes: 1) Guidelines development by such groups as the Outcome Measures in Rheumatology Clinical Trials (OMERACT) Group and the Osteoarthritis Research Society International (OARSI), and 2) the sudden demand for alternate-language translations of standardised, valid, reliable and responsive health status measurement tools to meet an international measurement need in OA for the evaluation of cyclo-oxygenase-2 (COX-2) inhibitor class compounds. These developments eventuated in the early to mid-1990's, and both the guidelines development by OMERACT and OARSI, and subsequently by regulatory agencies, as well as the ongoing need for standardised measures in clinical research environments continue to drive the globalisation of the WOMAC Index.

The guidelines for the testing of SADOA, derive from the recommendations of a WHO/ILAR Working Party, meeting in association with the newly formed Osteoarthritis Research Society. The resulting guidelines addressed various aspects of clinical trial design for this class of intervention (19). The guidelines identified five domains (pain, function, consumption of analgesics and/or NSAIDs, physical examination and quality of life) (19). These guidelines were particularly important since they recognised the importance of measuring physical function and quality of life, as well as pain, and provided support for use of the WOMAC Index in SADOA class trials (20), (see commentary by Professor J Edmonds in Summary and Conclusions on page 77 of The Journal of Rheumatology 1994; Supplement 41) (19).

The OMERACT III meeting held in Cairns, Australia, is of particular importance since it used evidence-driven, consensus-based decision-making to specify domains of measurement for future OA clinical trials of knee, hip and hand OA. The process involved the presentation of data on existing measurement alternatives (21), followed by discussions and polling sessions (22). The end result, with participation of academics, rheumatologists, measurement experts, pharmaceutical and device manufacturers and representative of regulatory agencies, was $\geq 90\%$ agreement on pain, physical function and patient global assessment, as core set clinical measures for all future OA knee, hip and hand studies (22). The OMERACT III consensus stopped short of specifying specific instruments for measuring the three core clinical domains, and, by agreement, this was postponed until the OARSI Task Force meeting scheduled for a few weeks later in Washington DC.

The OARSI Task Force Guidelines for the design and conduct of clinical trials in patients with OA also involved multiple stakeholders from different disciplines, and used a similar process to that employed by OMERACT to achieve evidence-driven, consensus-based decisions regarding clinical trials methods and outcome measurement procedures(23). The OARSI ratified the OMERACT core clinical set specification of outcome measurement domains, and also specified outcome measurement techniques and instruments, which included recognition of the WOMAC Index (23). Appendices to the main OARSI document dealt with special measurement topics including Appendix II, which described issues relating to clinical assessment techniques (23).

Most recently, an ostensibly North American consensus group called the Initiative on Methods, Measurement and Pain Assessment in Clinical Trials (IMMPACT) Group

have developed evidence-based consensus around outcome measurement domains (24) and instruments (25) for clinical trials in patients with chronic pain. The latter publication also recognizes the relevance of the WOMAC Index for trials involving patients with OA.

The OMERACT and OARSI guidelines, in particular, resulted in a rapid increase in demand for the WOMAC Index, not only in its original language form of English for Canada, but also in multiple other languages, and in both the 5-point Likert and 100 mm visual analogue scaling formats. The development of the alternate-language translations of the WOMAC 3.1 Index was undertaken under my copyright, by Health Outcomes Group in Palo Alto California, using their standard operating procedures (SOP). In summary, the SOP involved the following seven steps: a) tandem forward translation to target language (fluently bilingual translators), b) reconciliation of any differences between target language translations, c) tandem backward translation to English (fluently bilingual translators), d) reconciliation of any differences between source (English) and backward translations, e) agreement on a proposed forward translation, f) linguistic validation on-site in a small group of relevant individuals, and g) finalisation of the alternate-language form. The time frame for the WOMAC 3.1 Index was standardised at 48-hours, and patients were directed to think about a specific joint (either a hip or knee, depending on the study). In order to address diversity in the use of language, the instructions to patients were enhanced and some question stems expanded to include more prepositions and common language. The WOMAC 3.1 Index, thereafter, became the standard form of the WOMAC Index.

The rigorous development and validation of the original WOMAC Index, the state-of-the-art SOP by which the WOMAC alternate-language translations were created, the immediate availability of over 60 alternate translations of the WOMAC 3.1 Index in the end-users preferred scaling format (Likert or VA), and the capacity of the WOMAC Index to meet many of the requirements of recently introduced measurement guidelines, aligned with those of regulatory bodies or recognised by those bodies, undoubtedly contributed to the rapid globalisation of the WOMAC Index (26).

References:

19. Lequesne, M., Brandt, K., Bellamy, N., Moskowitz, R., Menkes, C.J., Pelletier, J-P. and Altman, R. Guidelines for testing slow acting drugs in osteoarthritis. Journal of Rheumatology 1994;(Suppl 41)21:65-73.

20. Bellamy, N. Outcome measurement in osteoarthritis clinical trials. Journal of Rheumatology 1995;(Suppl 43)22:49-51.

21. Bellamy, N. Osteoarthritis Clinical Trials: Candidate Variables and Clinimetric Properties. Journal of Rheumatology 1997;24:768-778.

22. Bellamy, N., Kirwan, J., Boers M, Brooks P, Strand V, Tugwell P, Altman, R., Brandt, K., Dougados, M., and Lequesne, M. Recommendations for a core set of

outcome measures for future phase III clinical trials in knee, hip and hand osteoarthritis. Consensus development in OMERACT III. Journal of Rheumatology. 1997;24:799-802.

23. Osteoarthritis Research Society (OARS) Task Force Report: Design and Conduct of Clinical Trials of Patients with Osteoarthritis: Recommendations from a Task Force of the Osteoarthritis Research Society. Osteoarthritis & Cartilage 1996;4:217-243.

24. Turk DC, Dworkin RH, Allen RR, Bellamy N et al. Core outcome domains for chronic pain clinical trials: IMMPACT Recommendations. Pain 2003;106(3):337-45.

25. Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen, Katz NP, Kerns RD, Stucki G, Allen RR, Bellamy N et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain 2005;113:9-19..

26. Bellamy N. WOMAC: A 20-year experiential review of a patient-centered self-reported health status questionnaire. Journal of Rheumatology 2002;29(12):2473-2476.

Guidelines for Testing Slow Acting Drugs in Osteoarthritis

MICHEL LEQUESNE, KENNETH BRANDT, NICHOLAS BELLAMY, ROLAND MOSKOWITZ, CHARLES JOEL MENKES, and JEAN-PIERRE PELLETIER

ABSTRACT. New compounds appear to improve symptoms of osteoarthritis (OA), and others are putative chondroprotective agents. We suggest experimental designs for studying the effects of these agents in subjects with hip and knee OA. The course of the articular cartilage lesion is the primary outcome measure to be assessed in putative chondroprotective agent trials. Serial radiographic studies suggest that the annual rate of joint space narrowing in patients with hip or knee OA is about 0.25 mm. Other approaches to quantitation of cartilage loss, e.g., radiographic measurement of the area of joint space, ultrasonography, magnetic resonance imaging and fiberoptic arthroscopy (for knee OA) are under investigation. (*J Rheumatol 1994;*(suppl 41)21:65–73)

Key Indexing Terms:
OSTEOARTHRITIS
CHONDROPROTECTIVE DRUGS
SLOW ACTING DRUGS IN OA
TRIAL METHODOLOGY

Two classes of slow acting drugs for the treatment of osteoarthritis (OA) — SADOA — will be considered: (1) Symptomatic Slow Acting Drugs for Treatment of Osteoarthritis (SYSADOA) and (2) Disease Modifying OA Drugs (DMOAD), sometimes inappropriately called chondroprotective agents. Both classes may be administred by the oral, parenteral, or intraarticular route.

Symptomatic slow acting drugs for treatment of osteoarthritis (SYSADOA). Recently, some drugs that are neither rapidly acting analgesics nor nonsteroidal antiinflammatory drugs (NSAID), nor chondroprotective (disease modifying) agents, but have a slow onset of action, have been alleged to improve OA symptoms. Their onset of action occurs only after a period of weeks, and symptomatic relief may continue for a considerable period after cessation of treatment. Some of these agents are administered orally or parenterally (e.g., chondroitin sulfate[1], glucosamine sulfate[2], diacerrhein[3]); others (e.g., hyaluronic acid[2,4], orgotein[2]) are administered by intraarticular injection. Although these agents have been studied in controlled clinical trials, additional studies are necessary to demonstrate whether any are efficacious.

To ascertain its efficacy, a SYSADOA should, in general, be compared against placebo, analgesics, or an NSAID of proven efficacy, using classical techniques in a trial of moderate length, i.e., several months to 1 year. Target populations might include subjects with either early or advanced OA. Outcome measures should, as in trials of NSAID, assess the effect of treatment on joint pain, function, and quality of life.

Disease modifying antiosteoarthritis drugs (DMOAD). The term chondroprotection introduces the concept that some drugs may slow the rate of articular cartilage degeneration, and/or enhance the rate of cartilage repair. Conversely, the term chondroaggression has been applied to suggest that some drugs (NSAID) may accelerate the rate of cartilage degeneration or inhibit cartilage repair. Both concepts derive primarily from results obtained with animal models and *in vitro* experiments, rather than from validated observations in humans.

Our present knowledge suggests that some abnormalities in OA cartilage may be reversible (e.g., proteoglycan depletion) and others irreversible (e.g., damage to the collagen network). These principles must be taken into account when clinical trials of DMOAD are designed and primary outcome measures are established.

Data from well designed clinical studies do not now exist to support the contention that any drug is chondroprotective, i.e., prevents, retards, or reverses cartilage lesions in humans. However, the search for agents capable of preventing OA progression and/or reversing established OA is strongly encouraged.

Patient groups suitable for study of a DMOAD. Depending upon the objective of treatment, the following groups of patients may be used to evaluate the efficacy of a DMOAD (Table 1).

Group 1. To assess the ability of the drug to prevent OA in an at-risk joint, subjects with an injury known to be associated with development of OA in a high proportion of

From Hôpital Léopold Bellan, Paris, France; Multipurpose Arthritis and Musculoskeletal Diseases Center, Indianapolis, IN, USA; University of Western Ontario, London, ON, Canada; Division of Rheumatology, University of Cleveland, Cleveland, OH, USA; Hôpital Cochin, Paris, France; Rheumatic Diseases Unit, and University of Montreal, Montreal, PQ, Canada.

M. Lequesne, MD, Professeur au Collège de Médecine des hôpitaux de Paris; K. Brandt, MD, Professor of Medicine; N. Bellamy, MD, Professor of Medicine and Epidemiology; R. Moskowitz, MD, Professor of Medicine; C.J. Menkes, MD, Professeur à l'Université Paris V; J.P. Pelletier, MD, Associate Professor of Medicine.

Address reprint requests to Pr. M. Lequesne, Hôpital Léopold Bellan, 19, rue Vercingétorix, 75014 Paris, France.

Table 1. *Requirements for clinical trials of SADOA*

	Trials of SYSADOA	Trials of DMOAD	
Aim	Improvement of symptoms	Slowing, arrest, or reversal of articular cartilage lesions	Prevention of articular cartilage lesions
Study population	Patients with painful OA of hip, knee, and (? hand)	1. Patients with painful idiopathic hip or knee OA without complete loss of articular cartilage (Grade II–III Kellgren-Lawrence, Grade I by chondroscopy) 2. Patients with OA of the hand	1. Victims of mechanical joint derangement* 2. Subjects with vocational or avocational risk for OA**
Main assessment criteria	Pain scales Algofunctional indices Consumption of analgesics and/or NSAID Physical examination Quality of life scale	Status of articular cartilage, based on radiography (joint space) or fiberoptic arthroscopy (knee) or, possibly MRI, ultrasonography, fiberoptic arthroscopy. Rate of change monitored on serial images of the affected joint.	Time interval between trauma or exposure and onset of OA
Duration of trial	3–12 months	2–4 years	Many years
Difficulty in recruitment	Minimal	Minor. Requires symptomatic patients with incomplete narrowing of joint space on radiograph	Maximal. Requires asymptomatic subjects undergoing longterm treatment of only theoretical benefit.

* e.g., anterior cruciate ligament rupture, meniscectomy, meniscus tear.
** e.g., football, judo, dance, rugby.

cases, e.g., anterior cruciate ligament rupture, would be suitable.

Group 2. To assess the ability of the drug to retard the progression of established OA, patients with symptomatic idiopathic OA of knee or hip with incomplete loss of joint space, or early hand OA, would be suitable.

Group 3. To assess the ability of the drug to repair damaged cartilage (disease reversal), patients with OA at any stage would be suitable.

Patients in Group 2 are the most likely targets for study.

Outcome measures in studies of DMOAD. For studies of these groups, outcome measures must be clearly defined, as follows.

Anatomical assessment of disease progression. Current methodologies continue to undergo validity and reliability testing. In prospective studies conducted over 3 years, measurements of the joint space, using special calipers on plain radiographs of the hip or knee of patients with OA, showed a mean rate of loss of about 0.25 mm[5–7] to 0.30 mm[8] per year, with a coefficient of variation (intraobserver reproducibility) of about 8–10%. Measurement of a relevant area of the joint space may be more useful than measurement of the width at a given point[9,10]. Other techniques, such as microfocal radiography, ultrasonography[11], magnetic resonance imaging (MRI), and computerized tomography, are under consideration. For evaluation of the knee joint, arthroscopy (currently the gold standard for pathologic findings) should be considered. Use of a small bore fiberoptic arthroscope is less invasive than standard surgical arthroscopy, and use of the needle arthroscope for serial evaluation of a patient is ethically acceptable.

Biochemical/immunochemical markers. Biochemical or immunochemical measurement of the concentration in synovial fluid, serum, or urine of indicators of degenerative and/or reparative responses in joint tissues is possible. However, while levels of these variables may reflect the magnitude of a biologic response in OA, none has been shown to reflect the adequacy of such responses within joint tissues. Before such measurements can be recommended for clinical use, further reliability and validity testing are required[12].

Since no reference drugs are available for use as standards for testing DMOAD, a placebo control group should be used. Assessment of the efficacy of a DMOAD will require years, rather than weeks or months.

Experimental prerequisites (DMOAD). The results of *in vitro*, *ex vivo*, or *in vivo* studies of the effects of drugs on cultures of articular cartilage or chondrocytes from animal models do not necessarily predict the results of clinical trials in humans with OA. It is essential, nonetheless, that evidence of a beneficial effect of a prospective DMOAD be obtained *in vivo* in animal models before trials in humans are undertaken.

Although, theoretically, a drug that is ineffective *in vitro* could be efficacious *in vivo*, and vice versa, it is reasonable to require that a potential DMOAD demonstrate activity *in vitro* and then *in vivo* in animal models of OA before a clinical trial in humans.

The most widely used *in vitro* methods for assessing the

effects of drugs on cartilage involve measurement of levels of degradative enzymes or the biosynthetic activity of the chondrocyte in cell cultures or organ cultures of animal or human OA cartilage[13,14]. The most widely used animal models of experimental OA involve anterior cruciate ligament transection in the dog (the Pond-Nuki model), partial medial meniscectomy in the rabbit (the Moskowitz model), and standardized contusion of the patella in the rabbit (the Mazieres model). Models using intraarticular injection of iodoacetate or papain have also been employed[15].

Among the many issues requiring consideration, determination of the dose of the DMOAD to be tested *in vivo* is one of the most difficult. This determination must be based upon the results of toxicology studies, which are essential before implementation of a clinical trial.

Demonstration that a candidate DMOAD possesses desirable properties *in vitro* (e.g., enhancement of chondrocyte proliferation, synthesis of type II collagen and proteoglycans, reduction of enzymatic degradation of the matrix) is of uncertain significance and does not necessarily represent chondroprotection. An effective DMOAD will, by definition, preserve or improve the morphology of hyaline articular cartilage in the OA joint.

Many *in vitro* or *ex vivo* studies of potential DMOAD have been conducted using normal articular cartilage. While important, this is only a first step, since the metabolic effect of a drug on OA cartilage may be different from its effect on normal cartilage.

Moreover, efficacy of a DMOAD should be demonstrable in at least 2 well established animal models of experimental OA, involving animals of 2 different species. Furthermore, consistency of *in vitro* and *in vivo* findings should be demonstrated.

It is desirable not only that results of *in vivo* testing in different experimental models of OA are consistent, but that the observations are confirmed by more than one investigator.

Unfortunately, to date, among those NSAID or SADOA that have shown a chondroprotective effect *in vitro* or *in vivo* in animals, none has proved chondroprotective in well controlled clinical trials in humans with OA.

Trials of SYSADOA in human OA. Aim. As with a rapidly acting drug, the aim is to reduce the severity of symptoms of OA in humans, and trials should be conducted in accordance with WHO/EULAR recommendations.

Criteria for assessment of efficacy. SYSADOA trials should take into account (1) the lag in the onset of improvement (induction–response interval) (2) the duration of the residual effect on symptoms after cessation of treatment; (3) changes in the requirement for analgesics or NSAID during the first weeks of treatment and later; (4) changes in joint mobility: while measurements of mobility are insensitive in short term studies, they may be of interest in trials of longer duration.

The investigator should choose a reasonable number of assessment measures (generally 5 to 8). One of these should be designated the primary outcome measure. The following should be considered: (1) Visual analog scale (VAS) for pain. This may be applied separately to different types of pain (e.g., nocturnal, weight bearing); (2) Functional indices. Two have been specifically designed and validated for OA: the severity, or algofunctional, indices for hip and knee OA[16] and the WOMAC Osteoarthritis Index for the same joints[17,18]; (3) Doyle Index[19] in its complete form or limited to specific joints of interest, e.g., hips, knees, or hand joints; (4) Loss of joint mobility (range of motion), although the reproducibility of this measure is not good[20]; (5) Walking time over 20 or 50 meters, or the time to go up and down a standard flight of stairs (applicable to hip and knee OA). In a multicenter trial, however, the stairs may not be sufficiently standardized; (6) Level of consumption of analgesics and/or NSAID; (7) Number of flares over time (especially with effusion in patients with OA of the knee); (8) Patient's overall (global) judgement of efficacy; (9) Investigator's overall (global) judgement of efficacy; and (10) Quality of life scale. A simple QOL scale is needed that is reliable, valid, and sufficiently sensitive to detect clinically important change in trials of SYSADOA.

As a criterion of efficacy, one of the above outcome measures must be chosen as primary *a priori*. This will help in calculating the sample size and will serve for primary statistical analysis. Moreover, although clinical endpoint measures have been validated in numerous short and moderate term trials, this is not yet the case for longterm trials (2–3 years). Good candidates are the algofunctional indices of OA severity (i.e., WOMAC and Lequesne's indices, and their components), and various other measures of pain, stiffness, and physical disability.

Physical and psychosocial adaptation may influence some of these endpoint measures, especially in longterm studies, and should be taken into consideration. However, the effect of this possible source of bias should be eliminated by the randomization process in randomized controlled trials.

Assessment of side effects. The occurrence of adverse reactions should be assessed by traditional methods. Followup must be continued for 2 to 3 months after completion of treatment, since the effect of SYSADOA is often longlasting.

Types of patients and eligibility criteria. OA of the hip and tibiofemoral OA are the best models. Patellofemoral joint OA, if this is the only, or the most prominent, site of OA in the knee, should be excluded, since it represents an entity distinct from tibiofemoral OA and symptoms are highly variable.

The diagnosis of OA must be confirmed according to defined criteria. Those quoted by the WHO/EULAR guidelines of 1986 were set up by Lequesne[21], those of Altman, *et al* (American College of Rheumatology) have recently been validated[22–24]. Patients with chondrocalcinosis and other

causes of secondary OA should be excluded. Exclusions for age, concomitant disease, and other variables should be appropriate.

The hip or knee joint should be painful daily, or at least on more than half of the days of the previous 2 months. Pain severity should be sufficient to permit detection of change (for example, 35 mm on a 100 mm VAS). The severity index[16] should usually be between 3 and 12 points.

Since the assessments will monitor only change in symptoms, the stage of OA is immaterial. However, patients with very far advanced OA (e.g., Kellgren-Lawrence Stage IV) should be excluded, as should those awaiting arthroplasty. Informed consent must be given by every patient.

Design and duration of the trial. The trial should be controlled, randomized, double blinded, and parallel in design. Since no reference SYSADOA exists, placebo should be used for the comparison group, with use of a rescue analgesic, if necessary. The duration of the trial should be at least 4–6 months. If a placebo group is not employed, a control group taking NSAID could be used alternatively. When feasible, both a placebo control group and an NSAID/analgesic group should be included in initial studies of these agents.

Concurrent treatment. Physiotherapy and occupational therapy may be continued if the patient has already been on a stable program and continues on the program throughout the trial. Intraarticular injections of any type should be prohibited for 3 months before the onset and throughout the trial.

An important consideration is the concomitant use of analgesics and NSAID. In most trials of SYSADOA, continued use of analgesics and/or NSAID has been permitted and the rate of consumption has been used as an outcome measure. However, some investigators prefer not to use reduction in NSAID intake as a primary criterion, and some even prohibit NSAID use during the study, using only a rescue analgesic, as needed. If that approach is used, the patient should be asked to terminate analgesic treatment at a sufficient interval before each assessment to wash out any analgesic effect. This method would focus the assessment on the effect of the SYSADOA.

Statistical analysis. The number of patients required must be calculated in advance. Withdrawals represent an important problem in moderate and longterm studies. An intention-to-treat design is preferable, with the final observation providing the endpoint for analysis. The method based on survival curves can solve the problem if outcome is categorized only as "success" or "failure."

Labelling. The label chondroprotective agent is not warranted for a SYSADOA unless it can be shown to predictably decrease the rate of cartilage loss. Similarly, the designation "antiosteoarthrotic substance" should be avoided unless the agent has been shown to retard progression of OA.

Trials of DMOAD (chondroprotective agents)

Definition of chondroprotection and objective of the trials. DMOAD (chondroprotective) therapy in OA, by definition, prevents, retards, or reverses the articular cartilage lesions in the disease *in vivo* in humans. To satisfy this definition, morphologic evidence of prevention, stabilization, or slowing of destructive lesions in the articular cartilage of the involved joint is required.

Criteria for assessment of efficacy. The primary outcome measure is the rate of articular cartilage loss over years, determined by radiography or some other method[5].

Radiographic measurement of the rate of joint space narrowing (JSN). (1) Measurement at the point of maximal narrowing. In patients with hip or knee OA, the rate of JSN may be measured on serial high quality radiographs of the pelvis or the knees taken in the standing position. Ideally, measurements should be made with a special caliper with 2 sharp points and a stabilizing screw. A magnifying glass with 0.1 mm graduation is used to measure the distance between the 2 sharp points. The joint space at the site of maximal narrowing may be measured with a precision of 0.1 mm[25]. With this manual technique, in a controlled, prospective, randomized, 3 year double blind trial comparing glycosaminoglycan peptide complex (Rumalon) and placebo in 42 patients with hip OA, Lequesne, *et al*[6] found a mean rate of JSN of 0.22 mm/year, with a very large inter-patient variation: 95% CI = 0 to 0.82 mm/year (Figure 1). Tran, *et al*[8], using computerized image analysis[9], found a mean loss of 0.30 mm/year.

Using the manual technique in a prospective study of patients with femorotibial OA, Lequesne, *et al* found a mean rate of medial compartment JSN of 0.24 mm/year[7] (Figure 2) in patients followed for nearly 4 years. In contrast, in a recent abstract, Kirwan, *et al*[26] reported that measurement of joint space width directly with a ruler on the radiograph showed a loss of 15% (= 0.6 mm/year) in one group and of 4% (= 0.2 mm/year) in another group of patients with knee OA. The marked discrepancy in the rate of JSN in these 2 studies is unexplained. (2) Measurement of the area of the

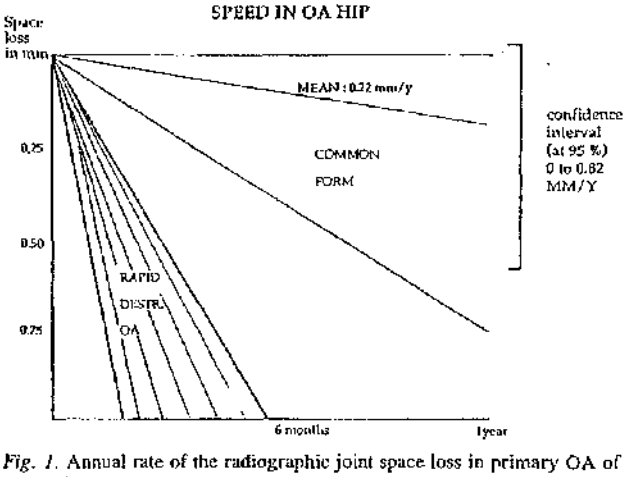


Fig. 1. Annual rate of the radiographic joint space loss in primary OA of the hip[6].

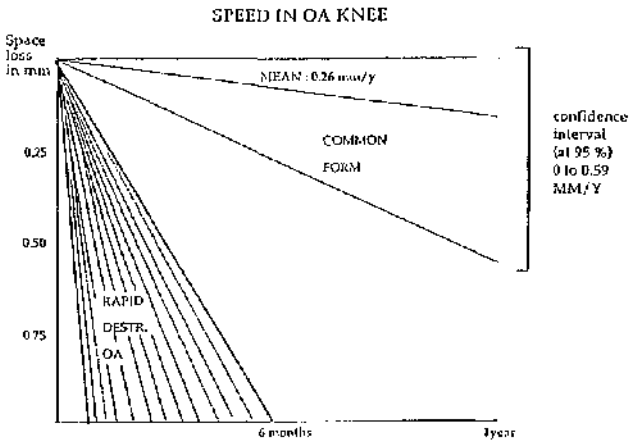


Fig. 2. Annual rate of the radiographic joint space loss in tibiofemoral OA of the knee[7].

joint space. Investigators have examined the possible advantages of measuring the area of the joint space on digitized radiographic images of the joint[9,10,26,27]. The technique has good reproducibility and may be preferable to direct measurement of the joint space since it encompasses the entire relevant area of the joint space, some part of which generally remains available for measurement over years. However logically, measurement of the JS area is less sensitive to change than that of the JS narrowest point as argued by Buckland-Wright in a remarkable review[28].

Obviously, the rate of progression of JSN in an individual patient, determined from the specific reduction of joint space on serial examinations, must take into account the sensitivity of the method employed (coefficient of variation 8–10%)[6,7]. (3) Validation of the joint space measurement method. In view of the length of the trial necessary to evaluate a DMOAD, the observation that serial radiographs may show progressive loss of joint space over years suggests that this measurement may provide a global view of articular cartilage thickness. It should be noted, however, that whether radiographic measurement of the joint space accurately reflects the thickness of cartilage in the knee or hip has not been confirmed by correlation with measurements on anatomical specimens. On the other hand, while newer techniques, such as ultrasonography[11], MRI, or arthroscopy[29], may provide reasonably accurate and sensitive measurements of articular cartilage thickness or lesions, they have not been employed in prospective longterm studies to determine an annual rate of cartilage loss. (4) Microfocal radiography. This technique, advocated by Buckland-Wright[28,30,31], provides higher resolution than conventional radiography, but the equipment is expensive and not generally available. The technique is, therefore, not well suited to a multiregional trial. (5) Measurements in patients with bilateral disease. In cases with asymptomatic contralateral involvement, progression of JSN should be measured bilaterally.

Other imaging techniques. In addition to the radiographic techniques described above, other imaging methods, e.g., ultrasonography and MRI, show promise for assessing the thickness of articular cartilage, but these remain to be validated in cohort followup studies.

Needle arthroscopy (Chondroscopy). This modified needle arthroscopic technique, currently useful only for the knee, is being validated for quantification of cartilage lesions[29,32]. It may be more sensitive than radiographic techniques and permit trials of shorter duration. The main limitations of this technique are inconsistency in assessing and mapping the chondral lesions and the ethical problem of repeated studies in patients who improve symptomatically with treatment. However, Ayral, *et al* have recently scored the lesions with an acceptable reproducibility, provided the chondroscopy is performed by the same well trained investigator[32].

Biochemical and immunochemical markers of OA. Serologic markers of cartilage breakdown or repair have not been shown to be of value in quantifying the progression of cartilage breakdown or the adequacy of cartilage repair in OA[12].

Clinical assessment. Tests for efficacy and safety of DMOAD. Since stabilization or improvement in the status of articular cartilage in an OA joint would be inconsequential unless accompanied by improvement in algofunctional status, a DMOAD should be assessed also by the clinical outcome measures for evaluation of a SYSADOA described above, measures of importance since symptoms and radiographs often did not correlate[33,34].

Eligibility and exclusion criteria. Two types of individuals are potentially suitable for DMOAD trials: (1) those at risk for future OA because of previous trauma, malformation, or occupation, in whom preventive treatment can be undertaken; (2) those with symptomatic hip and knee OA, in whom maintenance of radiographic joint space width over several years will be the primary assessment criterion. Consequently, JSN at the onset of the trial should be incomplete.

Patients under 50 years of age should be excluded, since OA tends to progress slowly in younger individuals.

Patients with complete congenital dysplasia and dislocation of the hip, OA secondary to acetabular protrusion, osteonecrosis, previous articular fracture, or significant genu varum should also be excluded, since the course of the disease is often more rapid in these conditions.

OA of the interphalangeal joints of the hand could be used as an alternative model. Outcome measurements would include assessment of the course in affected joints (despite obvious imaging difficulties[10,30]) and the development of OA in previously unaffected joints.

As a general principle, in trials involving patients with hip

OA or knee OA it is important also to evaluate other sites of potential or actual involvement, e.g., interphalangeal joints.

Design and duration of the trial. In all cases, the trial should be controlled, randomized, double blinded. A parallel design should be used and, until a reference drug is identified, should include a placebo group. It should be conducted over an appropriate number of years in well characterized patient groups that are as homogeneous as possible with respect to disease characteristics.

In trials of agents in which the aim is prevention of OA, analysis of the results is based on survival curves. Only one outcome criterion is relevant: development or absence of OA at a specified location, as determined by periodic radiographic or other imaging techniques. The curve might look like the theoretical one in Figure 3.

However, a trial aimed at demonstrating a prophylactic effect is rather hypothetical and would be difficult to implement. The duration of the trial would be unknown, since the rate of occurrence of OA is extremely variable (and is often measured in decades, even in subjects predisposed to OA, e.g., those with cruciate ligament transection). Furthermore, recruitment of a sufficient number of subjects who would be prepared to be followed over years on a treatment of uncertain value would be difficult, particularly if none is symptomatic and 50% are assigned to the placebo group.

In trials of agents in which the aim is the stabilization or repair of established OA, given that the rate of articular cartilage loss is usually slow (Figures 1, 2), the duration of the trial should be no less than 2 to 3 years.

Concomitant therapy. Concomitant treatment that is potentially chondroprotective or chondroaggressive should obviously be prohibited during a study involving a putative DMOAD. In a longterm trial it is desirable to permit only analgesics. NSAID should be used only if essential, for short periods, as a rescue medication.

The change in consumption of analgesics, reflecting the level of joint pain, may be a useful secondary clinical assessment measure, if analgesic use is strictly monitored.

Intraarticular injection of the target joint should be prohibited.

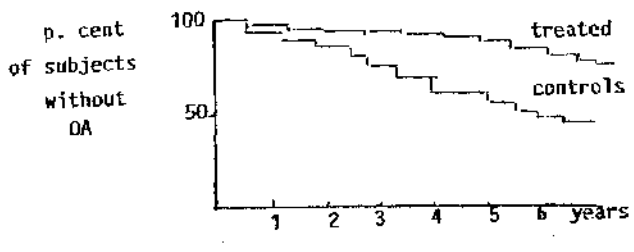


Fig. 3. Theoretical example of a positive effect on a survival curve regarding the prevention of OA by a putative DMOAD in at-risk subjects.

Statistical analysis and calculation of sample size. Statistical analyses can be employed using techniques suitable for comparing the experimental and control groups. If radiographic evidence of JSN is employed as the primary outcome measure, the sample size calculation should be based on studies that have measured the rate of joint space loss[6,8]. For example, in patients with OA of the hip, the mean loss over a 3 year period was 0.66 mm, with a large SD of ± 0.93 mm[6]. Consequently, to obtain a difference between the treated group and the placebo group of 0.5 mm in joint space loss over a 3 year period, with type I error = 0.05 and type II error = 0.10 in a 2-tailed test, it would be necessary that 163 patients per group complete 3 years in the study, and, for a gain of 0.33 mm within 3 years, 86 patients per group. The coefficient of variation of the measurement (8–10%) was taken into account in our calculation.

In order not to lose any information from patients who withdraw from treatment, an analysis based on survival curves should be performed. This offers the considerable advantage of accounting for all patients who are enrolled. Those who withdraw due to inefficacy or adverse reactions can be analyzed separately or together. A theoretical example of such a curve is shown in Figure 4.

Labelling. Unless evidence of true protection of human OA cartilage has been demonstrated, the designation "chondroprotective drug" should not be used.

Conclusion. Although a number of putative SYSADOA and DMOAD are already available for clinical use in certain countries, it is essential that such agents undergo further rigorous evaluation. In this article, which was written under the authority of the International League of Associations for Rheumatology (ILAR), we have tried to summarize the relevant methodology for study of both classes of SADOA.

Two important issues remain unresolved: (1) For clinical assessments, what is the best primary criterion of efficacy (i.e., outcome measure)? Will the principal outcome measures prove to be satisfactory in longterm trials? And, especially in trials of SYSADOA, should concomitant treatment with an NSAID be permitted, so that possible reduction of NSAID intake may serve as an outcome measure? It would

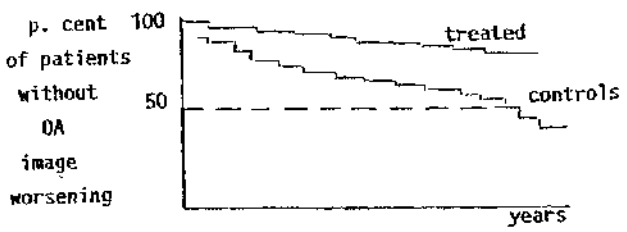


Fig. 4. Theoretical survival curve showing a positive result of a DMOAD candidate to reduce the rate of patients with OA with worsening of their established OA image.

seem desirable that some trials of SYSADOA are conducted with the concomitant NSAID treatment and others without it. (2) With regard to studies of DMOAD, several methods for measuring articular cartilage loss over years are under evaluation. Presently, radiographic measurement of the rate of JSN is the simplest among these, and in longterm followup of patients with hip and knee OA has shown a consistent rate of loss. However, correlations of joint space width with actual measurements of cartilage thickness are not currently available.

REFERENCES

1. Mazières B, Loyau G, Menkes CJ, *et al:* Le chondroïtine sulfate dans le traitement de la gonarthrose et de la coxarthrose. *Rev Rhum Mal Osteoartic 1992;*59:466–71.
2. Noack W, Fischer M, Förster KK, Rovati L, Setnikar I: Glucosamine sulfate in osteoarthritis of the knee. *Osteoarthritis Cartil 1994;*2:51–9.
3. Nguyen M, Dougados M, Berdah L, Amor B: Diacerhein in the treatment of osteoarthritis of the hip. *Arthritis Rheum 1994;*37:529–36.
4. Peyron JG: A new approach to the treatment of osteoarthritis: Viscosupplementation. *Osteoarth Cartil 1993;*1:85–7.
5. Lequesne M, Lamotte J, Samson M: Conditions de démonstration radio-cliniques d'un effet chondroprotecteur ou chondroagresseur. In: Gaucher A, Netter P, Pourrel J, Regent D, eds. *Actualités en physiologie et pharmacologie articulaires. Paris: Masson, 1991:390–401.*
6. Lequesne M, Winkler P, Rodriguez P, Rahlfs VW: Joint space narrowing in primary osteoarthritis of the hip. Results of a three year controlled trial. *Arthritis Rheum 1992;*(suppl 9)35:S135.
7. Lequesne M, Orvain J, Massé JP, Gilmet T: Rate of the joint space narrowing (JSN) over 3–5 years in medial tibio-femoral osteoarthritis of the knee. *Rev Esp Reumatol 1993;*(suppl 1)20:28.
8. Tran AM, Conrozier T, Balblanc JC, *et al:* Etude de la vitesse du pincement de l'interligne articulaire dans la coxarthrose. *Rev Rhum Mal Osteoartic 1993;*60:385.
9. Conrozier T, Tran AM, Balblanc JC, *et al:* Measurement of the hip joint space using computerized image analysis. *Rev Rhum Mal Osteoartic 1993;*(Engl ed)60:105–11.
10. Byrne J, Dacre JE: Digital image analysis of the hands (DIP, PIP, MCP), knees and hips: A rapid and reproducible method for joint space measurement. *Arthritis Rheum 1992;*(suppl 9)35:S134.
11. Brandt K, Myers S, Dines K, Brandt D: High frequency ultrasound scans characterize normal (NAC) and osteoarthritis (OAC) articular cartilage. *Osteoarth Cartil 1993;*1:23.
12. Brandt KD: A pessimistic view of serologic markers for diagnosis and management of OA. Biochemical, immunologic and clinicopathologic barriers. *J Rheumatol 1989;*18:39–42.
13. Burkhardt D, Ghosh P: Laboratory evaluation of anti arthritic drugs as potential chondroprotective agents. *Semin Arthritis Rheum 1987;*(suppl 1)17:3–34.
14. Mankin HJ, Brandt KD: Biochemistry and metabolism of articular cartilage in osteoarthritis. In: Moskowitz RW, Howel DS, Goldberg VM, Mankin HJ, eds. *Osteoarthritis,* 2nd ed. Philadelphia: WB Saunders, 1992:109–54.
15. Moskowitz RW, Experimental models of osteoarthritis. In: Moskowitz RW, Howell DS, Goldberg VM, Mankin HJ, eds. *Osteoarthritis,* 2nd ed. Philadelphia: WB Saunders, 1992:213–32.
16. Lequesne M, Mery C, Samson M, Gerard DP: Indices of severity for osteoarthritis of the hip and knee. *Scand J Rheumatol 1987;*(suppl 65)18:85–9.
17. Bellamy N, Buchanan WW, Goldsmith H, *et al:* Validation study of WOMAC: A health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol 1988;*15:1833–40.
18. Bellamy N, Buchanan WW, Goldsmith H, *et al:* Validation study of WOMAC: A health status instrument for measuring clinically important patient relevant outcomes following total hip and knee arthroplasty. *J Orthop Rheumatol 1988;*1:95–108.
19. Doyle DV, Dieppe PA, Scotte J, *et al:* An articular index for the assessment of osteoarthritis. *Ann Rheum Dis 1981;*40:75–8.
20. Cushnaghan J, Cooper C, Dieppe P, *et al:* Clinical assessment of osteoarthritis of the knee. *Ann Rheum Dis 1990;*49:768–70.
21. Lequesne M: Clinical features, diagnostic criteria, functional assessments and radiological classification of osteoarthritis (excluding the spine). In: Huskisson EC, Katona G, eds. *Rheumatology 7. New trends in osteoarthritis.* Basel: Karger, 1982:1–10.
22. Altman R, Alargon G, Appelrouth D, *et al:* The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. *Arthritis Rheum 1991;*34:505–14.
23. Altman R, Alargon G, Appelrouth D, *et al:* The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hand. *Arthritis Rheum 1990;*33:1601–10.
24. Altman R, Asch E, Bloch E, *et al:* Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. *Arthritis Rheum 1986;*29:1039–49.
25. Laoussadi S, Menkes CJ: Amélioration de la précision de la mesure visuelle de la hauteur de l'interligne articulaire du genou et de la hanche à l'aide d'une loupe graduée. *Rev Rhum Mal Osteoartic 1991;*58:678.
26. Kirwan JR, Cushnagan J, Dacre J, *et al:* Progression of joint space narrowing in knee OA (abstr). *Arthritis Rheum 1992;*(suppl)35:B103,
27. Dacre JE, Huskisson EC: The automatic assessment of knee radiographs in osteoarthritis using digital image analysis. *Br J Rheumatol 1989;*28:506–10.
28. Buckland-Wright JC: Quantitative radiography of osteoarthritis. *Ann Rheum Dis 1994;*53:268–75.
29. Dedrick DK, Flechtenmache J, Chuta AS, *et al:* Detection of osteoarthritis of the knee using magnetic resonance imaging, needle and standard arthroscopy with histologic correlation. *Arthritis Rheum 1992;*35:S134.
30. Buckland-Wright JC, MacFarlane DG, Lynch JA, Clark B: Quantitative micro focal radiographic assessment of progression in osteoarthritis of the hand. *Arthritis Rheum 1990;*33:57–65.
31. Buckland-Wright JC, MacFarlane DG, Jasani MK, Lynch JA: Quantitative micro focal radiography detects changes in osteoarthritis knee joint space width in patients on diclofenac sodium versus placebo. *Osteoarth Cartil 1993;*1:23.
32. Ayral X, Dougados M, Listrat V, *et al:* Chondroscopy: A new method for scoring chondropathy. *Arthritis Rheum 1992;*35:S91.
33. Massardo L, Wadt I, Cushnaghan J, Dieppe P: Osteoarthritis of the knee joint: An eight year prospective study. *Ann Rheum Dis 1989;*48:893–7.
34. Dougados M, Gueguen A, Nguyen M, *et al:* Longitudinal radiologic evaluation of osteoarthritis of the knee. *J Rheumatol 1992;*19:378–84.
35. Lequesne M: Quantitative measurements of joint space during progression of osteoarthritis: chondrometry. In: Kuettner KE, Goldberg VM, eds. *New Horizons in Osteoarthritis,* Rosemont: American Academy of Orthopedic Surgeons 1995 (in press).

Outcome Measurement in Osteoarthritis Clinical Trials

NICHOLAS BELLAMY

N. Bellamy

ABSTRACT. The clinical assessment of outcome in osteoarthritis (OA) clinical trials is highly dependent on the use of valid, reliable, and responsive measurement techniques. Despite several decades of clinical studies, and a half-century of development in clinical metrology, we still lack international standards of measurement for OA trials. There have, nevertheless, been several very encouraging developments. In particular, the Osteoarthritis Research Society and the 5th WHO/ILAR Task Force have discussed issues of standardization. The Western Ontario and McMaster Universities Osteoarthritis Index and Lequesne Index have been proposed as important outcome measures. Finally, data have recently been published on observer variability, variance estimation, and sample size determination for OA trials. (*J Rheumatol 1995;*(**suppl 43)22:49–51**)

Key Indexing Terms:
OSTEOARTHRITIS OUTCOME MEASUREMENT CLINICAL TRIALS

The use of valid, reliable, and responsive measures is essential for the successful detection of clinically important alterations in health status in osteoarthritis (OA) clinical trials. During the last 50 years of musculoskeletal clinical metrology, numerous ad hoc methods of measurement have been employed. In contrast, several validated instruments, largely based on patients with rheumatoid arthritis (RA), e.g., Health Assessment Questionniare (HAQ) and Arthritis Impact Measurement Scales (AIMS), have been applied in OA studies. In addition, the Ritchie Index has been modified (i.e., Doyle Index) for use in patients with OA. Most recently there have emerged 2 purpose-built instruments for use in OA studies [i.e., The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) and the Lequesne Index].

Despite these developments, there has been relatively little standardization in outcome measurement in OA studies. Two recent reviews of published clinical trials suggest that there has been a high degree of variability not only in the variables selected but also in the instruments and scales employed for measurement. Measures of pain, patient global assessment, physician global assessment, and stiffness have been used in $\geq$ 50% of nonsteroidal antiinflammatory drug (NSAID) trials[1]. However, despite the clinical importance of physical disability, validated questionnaires have been infrequently employed in its measurement. While the 1985 EULAR guidelines[2] for OA clinical trials recommend measurement of physical disability by questionnaire, no specific mention of a similar measurement is made in the 1988 US Food and Drug Administration (FDA) guidelines[3]. In contrast, walk time is cited in both the EULAR and FDA guidelines, yet this measure seemed to lack responsiveness in the aforementioned clinical trials that were reviewed. Current FDA and EULAR guidelines differ in their listings of preferred outcome measures (Table 1). With few exceptions, insufficient detail is given in either set of guidelines regarding instruments, scales, or recording methods. The guidelines are not in agreement, one with the other, and both contain redundancies.

Most recently a set of guidelines for outcome measurement in trials of so-called slow acting drugs in OA (SADOA) has been proposed (Table 1)[4]. These guidelines contain 9 disease specific measures and one generic (quality of life) health status measure. They include measures of pain, physical function, physician and patient global assessments similar to the core measures developed by the Outcome Measures in Rheumatoid Arthritis Clinical Trials (OMERACT) group for use in RA trials. The relative merits of different articular indices (e.g., graded versus binary) have not been studied in OA, and, therefore, the role of the Doyle Index (partial or complete) remains to be defined. Overall, this proposal forms an excellent basis around which to discuss future instrumentation in SADOA trials.

Pari passu with attempts to develop guidelines for trials, there have been significant developments in the basic clinical metrology of OA. For example, it appears that patients with OA may rate their pain qualitatively in a different fashion than those with fibromyalgia or RA. Furthermore, pain varies and distinct circadian (acrophase = 19:20) and circaseptan (acrophase = Sunday) patterns of pain in patients with knee OA have been reported[5]. This suggests that the specific instrument used to measure pain in OA needs careful consideration, as does the time at which the measurement is made.

Our own index, the WOMAC Osteoarthritis Index, has been subject to 2 major validation studies and 16 additional investigations. We have compared Likert and visual analog versions of the index, blind versus informed presentations, parametric versus nonparametric forms of analysis, signal versus aggregate methods of measurement, the time frame dependency

From the Department of Medicine, The University of Western Ontario, London, ON, Canada.

N. Bellamy, MD, MSc, FRCP(Glas, Edin), FRCPC, FACP, Professor of Medicine, Epidemiology and Biostatistics, University of Western Ontario.

Address reprint requests to Dr. N. Bellamy, Suite 303, Colborne Building 3, Victoria Hospital, 375 South Street, London, ON, Canada N6A 4G5.

Table 1. *Guidelines for OA clinical trials*

FDA[3]:
1. Swelling
2. Redness
3. Tenderness on pressure
4. Pain at rest or on motion
5. Range of motion
6. Walking or stair climbing time
7. Investigator's opinion of the patient's condition on the day of assessment
8. Patient's opinion of his/her condition on the day of assessment

EULAR[2]:
1. Index of severity of hip and knee disease (Lequesne, 1982)
2. Investigator's overall opinion
3. Pain on visual analogue scale
4. Patient's overall opinion
5. Walking time (if knee, stair climb is recommended)

SADOA[4]:
1. Visual analogue scale of pain
2. Functional indices (e.g., WOMAC or Lequesne)
3. Doyle Index
4. Loss of mobility
5. Walking or stair climbing time
6. Consumption of analgesics and/or NSAID
7. Number of flares over time, especially effusion in OA knee
8. Patient's overall judgment of efficacy
9. Investigator's overall judgment of efficacy
10. Quality of life scale

of the response, and the importance of different index items, as well as the relative importance of scores on the 3 subscales to individual patients. We have also compared the relative statistical efficiency of WOMAC against the Lequesne, HAQ, AIMS, and Doyle indices, and against the 50 foot walk time, intermalleolar straddle (hip), intercondylar distance (hip), and range of movement (knee). These investigations suggest that WOMAC is, in general, slightly greater in statistical efficiency than other outcome measurement procedures. Most recently, we compared a computerized version of WOMAC against the original paper version, and have shown high levels of correlation between the 2 versions. WOMAC has been requested for translation into French, German, Italian, Spanish, Dutch, and Swedish, and validation of these foreign language translations is pending. To date, WOMAC has been requested for use by over 100 different investigators in 14 different countries.

In general, sample size calculation for clinical trials is based on a single variable. If this variable is pain, then the trial becomes a test of comparative analgesia. However, researchers are often interested in other outcomes. Our own research in this area, based on an analysis of NSAID trials and a series of studies examining variables for sample size calculation, suggests that sample size requirements differ depending on the variable employed[6-8]. Thus, with standardized observers and techniques, and assuming a 2-tailed $\alpha 0.05$ and $\beta 0.1$, the sample size estimates (per group) for a double blind randomized controlled parallel trial comparing 2 NSAID were as follows: Investigator's opinion of the patient's condition on the day of assessment = 25; patient's opinion of his/her condition on the day of assessment = 26; range of motion measurements = 29-31; pain on movement = 41; pain on visual analog scale = 47; tenderness on pressure = 76; 50 foot walk time = 94; and pain at rest = 119. Such differences in sample size requirements need to be acknowledged and accommodated in future trials. This in itself creates logistic problems. For example, the use of multiple variables may necessitate a statistical correction for multiple comparisons. Alternatively, weighting and aggregating information from multiple variables into a single value avoids the aforementioned correction, but is itself problematic. Nevertheless, the future success of OA trials depends on addressing, rather than ignoring, these issues. An alternative approach to aggregation is the development of response criteria, i.e., a combination of change values exceeding a specified magnitude on one or more predefined variables. This approach has been attempted by RA but not by OA researchers. Once a core set of variables is declared, the development of response criteria will naturally follow.

Much research to date has been based on disease specific measures, and yet there is an increasing interest in generic measures. Such measures offer an opportunity to compare different disease states and are a prerequisite for certain types of economic analysis. The use of measures such as the SF36 (The Medical Outcomes Study 36-Item Short-form Health Survey) and the Health Utilities Index (HUI) in OA studies has yet to be established, as has the sample size consequences of their inclusion. A recent presentation at the OMERACT II Conference, in Ottawa, suggested that in a total joint arthroplasty study, the SF36 was more responsive as a generic measure, but WOMAC was superior as a disease specific measure. In essence, both measures are excellent, but examine different aspects of the response to treatment.

Further research is required to rationalize the conceptual basis for measurement in OA studies. A framework that encompasses the symptomatology of OA and its consequences is illustrated in Figure 1. In essence the cellular and biochemical pathology of OA may result in a number of consequences that may be detected by imaging techniques such as plain radiographs, MRI, or chondroscopy. Ultimately, the disease may be clinically manifest and result in joint tenderness (Doyle Index), performance decrements (range of movement, 50 foot walk time) and the cardinal features of pain, stiffness, and physical dysfunction (WOMAC Osteoarthritis Index). If sufficiently severe, these and other associated effects may result in a decline in quality of life or overall health status (SF36 or HUI). In making global assessments, the patient and physician may independently take into account a variety of clinical consequences of OA. The physician may, in addition, consider the severity of disease detected by ancillary tests (e.g., radiographs, MRI, chondroscopy). By using

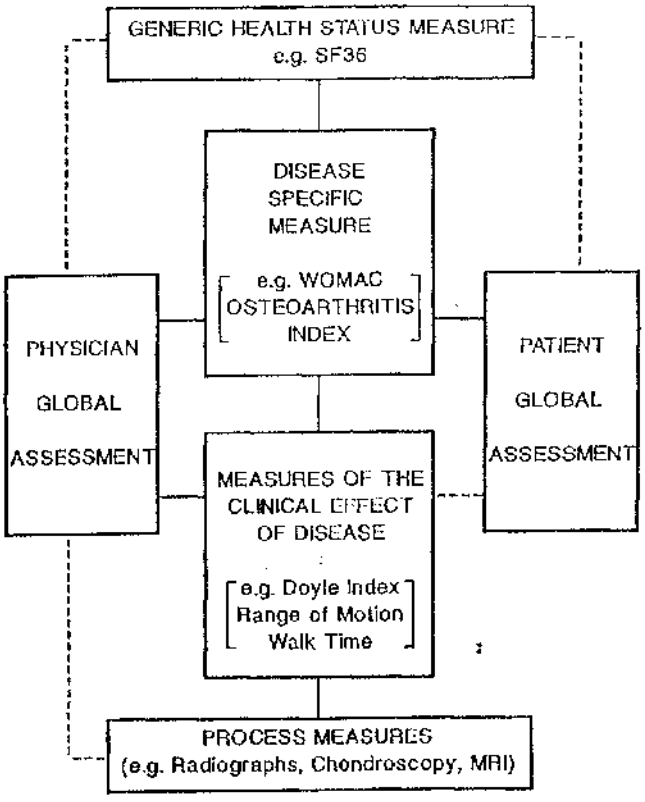


Fig. 1. A conceptual framework for outcome measurement in clinical trials of OA of the hip and/or knee.

a conceptual framework we can define not only what change is occurring, but where it occurs within the dimensionality of the symptomatology of the disease.

In summary, future needs in outcome measurement can be met by the simultaneous pursuit of several goals: (1) The encouragement of dialogue between various agencies that have developed guidelines to establish a truly international standard for studies having similar goals. (2) The preferential use of valid, reliable, and responsive outcome measures for future OA studies. (3) The routine incorporation of measures of physical disability in the outcome measurement process. For hip and knee studies this would entail the use of the WOMAC or Lequesne indices, and for generalized OA trials the HAQ or AIMS indices. (4) The further refinement of parameters used in sample size calculation. (5) Evaluation of the role of generic health status measures such as the SF36 and HUI in OA clinical trials. (6) OA researchers should follow the example of the OMERACT group in RA and develop a core set of measures for OA clinical trials and a set of response criteria to adjudicate the success versus failure of therapy in individual patients.

REFERENCES

1. Bellamy N: Osteoarthritis: In: *Musculoskeletal Clinical Metrology*. Dordrecht: Kluwer Academic Publishers, 1993:169–76.
2. Guidelines for the Clinical Investigation of Drugs Used in Rheumatic Diseases, European Drug Guidelines Series 5, World Health Organization, Regional Office for Europe, Copenhagen, European League Against Rheumatism, March 1985, 21–4.
3. US Department of Health and Human Services, Public Health Service, Food and Drug Administration: *Guidelines for the Clinical Evaluation of Anti-inflammatory and Anti-rheumatic Drugs (Adults and Children)*. April 1988, 12–5.
4. Lequesne M, Brandt K, Bellamy N, *et al*: Guidelines for testing slow acting drugs in osteoarthritis. *J Rheumatol 1994;* (suppl 41)21:65–73.
5. Bellamy N, Sothern RB, Campbell J: Rhythmic variations in pain perception in osteoarthritis of the knee. *J Rheumatol 1990;*17:364–72.
6. Bellamy N, Carette S, Ford PM, *et al*: Osteoarthritis antirheumatic drug trials. I. Effects of standardization procedures on observer dependent outcome measures. *J Rheumatol 1992;*19:436–43.
7. Bellamy N, Carette S, Ford PM, *et al*: Osteoarthritis antirheumatic drug trials. II. Tables for calculating sample size for clinical trials. *J Rheumatol 1992;*19:444–50.
8. Bellamy N, Carette S, Ford PM, *et al*: Osteoarthritis antirheumatic drug trials. III. Setting the delta for clinical trials — Results of a consensus development (Delphi) exercise. *J Rheumatol 1992;*19:451–7.

Osteoarthritis Clinical Trials: Candidate Variables and Clinimetric Properties

NICHOLAS BELLAMY

ABSTRACT. Outcome assessment in osteoarthritis (OA) clinical trials requires the use of valid, reliable, and responsive measurement techniques. Clinical assessment procedures examine patient relevant issues, while imaging procedures and biologic marker assays probe different aspects of the underlying disease process. At present, there is no international agreement on which standardized procedures should be used in future studies or which domains should be included. Nevertheless, significant progress has been made in recent years toward attaining that goal and there are numerous measurement procedures, differing in their conceptual and clinimetric properties, from which to select a core set. The following review provides a description of candidate variables that merit consideration in reaching international harmonization on outcome measures for future OA clinical trials. (*J Rheumatol 1997;24:768–78*)

Key Indexing Terms:
OSTEOARTHRITIS CLINICAL TRIALS OUTCOME

Osteoarthritis (OA) is often considered to be either of known (secondary) or unknown (primary or idiopathic) origin. Like most other organs, the clinical repertoire of the musculoskeletal system is relatively restricted. The major features of osteoarthritic involvement of a joint are pain, tenderness, swelling, crepitus, and dysfunction. Heat occurs occasionally, while redness is rare, except in cases of erosive arthritis and in the early stages of nodal OA in some patients.

The clinical metrology of OA is complex because, like conditions such as rheumatoid arthritis (RA), ankylosing spondylitis, and fibromyalgia, there are few constants in the clinical presentation. Furthermore, OA may be symptomatic or asymptomatic, and the associated radiographs either normal or abnormal. Indeed, the association between clinical and radiographic features is often loose, and there may be disparity between radiographic and arthroscopic aspects of the disease. Finally, the biochemistry of OA is complex and, despite considerable progress, current understanding is incomplete, there being no serologic test for OA and no biologic marker universally acknowledged as a marker of "disease activity" or of prognostic value.

For metrologic purposes, it is necessary to subcategorize OA. Only primary forms of the disorder will be further considered here, since patients with secondary forms of OA are usually excluded from most clinical trials. However, it must be acknowledged that ultimately the etiology of all forms of OA may be known, and that the disorder currently termed "primary OA" is probably a heterogeneous group of disorders varying in etiology. From a metrologic standpoint, primary OA can be divided into 7 categories: hip, knee, hand, and other forms of localized OA, and generalized, apophyseal, and erosive OA. I discuss hip, knee, hand, and generalized OA since trials of pharmacological agents are usually conducted in these areas. The importance of, for example, shoulder or ankle OA, or the pain and disability that may result from apophyseal or erosive OA should not be underestimated. However, the hands, hips, and knees are the most frequent target areas for primary OA and may be involved in a localized or generalized fashion. Furthermore, many measures suitable for assessing knee or hip OA are not appropriate for hand OA and are inadequate by themselves to assess patients with generalized OA. This difficulty of measuring different clinical configurations of the disease, from a metrologic standpoint, differentiates OA from RA.

From the University of Western Ontario, London, Canada.

N. Bellamy, MD, MSc, FRCP(Glas, Edin), FRCPC, FACP, Professor of Medicine, Epidemiology and Biostatistics.

Address reprint requests to Dr. N. Bellamy, Ste. 303, Colborne Bldg., London Health Sciences Centre, 375 South St., London, Ontario N6A 4G5, Canada.

CLINICAL VARIABLES

Before considering candidate variables for outcome measurement in future OA trials, it is necessary to consider a conceptual framework for conducting measurement in such studies[1]. Although linkages between the different consequences of disease may be rather loose, it is, pragmatically speaking, reasonable to consider a series of consequences arising from the underlying cellular pathology. In particular, clinical pathology leads to various clinical manifestations, and thereafter to a series of clinical outcomes (impairment, disability, handicap, even death). Conceptually, one can consider the existence of corresponding measurement strata. Depending on the disease and the dimension of interest, one could use (or develop) one or several appropriate instruments to probe each of the strata, or as many as are relevant given the research question and the known or predicted pharmacodynamic properties of the test compounds. Since the framework is presented as a hierarchical structure, it would be necessary to monitor all strata between the lowest

and highest stratum selected for inclusion. Indeed, it might be reasonable in some studies to monitor all clinical strata, since not only between-drug within-strata comparisons are relevant, but so are comparisons of the extent to which the treatment effects penetrate the hierarchical structure of disease consequence. Monitoring all strata, for example, might allow differentiation of a systemic corticosteroid from a nonsteroidal antiinflammatory type drug (NSAID). Similarly, a drug that improved functional status merely by elevating mood (and having little or no effect on lower strata) would have a different "therapeutic impact profile" from a drug whose effects were entirely mediated through its ability directly to affect the basic inflammatory process and any potentially reversible consequences. Thus, instead of categorizing outcome measures in order of alphabetic priority, degree of observer dependency or independency, primary or secondary outcome measurement status, or even random order, the measurement process could be described in terms of the strata monitored and the instrument(s) selected to monitor effects within each corresponding stratum. Such organization would allow a better understanding of where the primary outcome measure(s) fit(s) into the overall measurement framework. From the above construct, a number of candidate variables and measurement procedures for evaluating patients can be developed. These are identified in Table 1 and discussed in the following paragraphs.

Pain. The pain literature is extensive and diverse[2]. Although not developed specifically for or validated exclusively in patients with OA, the performance characteristics of both Likert and visual analog scales (VAS) have been evaluated in patients with OA[2]. Variations of these basic scales, such as the pain faces scale[3], continuous chromatic analog scale[4], as well as the McGill Pain Questionnaire[5], and Behavioural Observation Techniques[6,7], have not usually been used in evaluative trials in OA. As a result, their applicability and performance characteristics remain uncertain.

Experience with both 5 point Likert and 10 cm horizontal VAS in OA trials indicates they are both valid, reliable, and responsive techniques for assessing pain[8-10]. Since pain is a personal experience, pain scales should be completed by patients and not by independent assessors.

Stiffness. Because stiffness is generally of shorter duration in OA than untreated RA, it is often assumed to lack importance. However, patients asked to rate the absolute importance of stiffness indicate that it is of moderate importance[11]. Furthermore, patients asked to rate the relative importance of stiffness (cf. pain and physical function) assign importance scores of 21% in knee[12] and 25% in hand OA. Given its relatively short duration, it is preferable to assess severity rather than duration in patients with OA (cf. patients with RA). Valid, reliable, and responsive techniques based on 5 point Likert and 10 cm VAS have been developed for this purpose[8-10].

50 foot walk time. This measure is only applicable to the study of lower extremity involvement and is, therefore, inappropriate for assessment of hand OA or in isolation for the assessment of generalized OA. Even in the study of hip OA or knee OA patients, this measure has often been poorly responsive[13]. Despite being valid and reliable, it is questionable whether the 50 foot walk time is relevant or necessary[8-10].

Ascent time. Like the 50 foot walk time, this test is only applicable to lower extremity assessment. There is no definition of a standard set of stairs on which to perform the test, although the same set could be used for serial assessments of individuals. Performance on this test may depend on a number of factors including physical conditioning. Furthermore, comorbidities such as cardiovascular disease, neuropathy, and cerebellar dysfunction may result in this form of assessment being hazardous, particularly if the stairs are steep. I do not feel this measure is relevant for evaluative research in clinical trials.

Knee range of motion (ROM). This measure is only applicable to the study of knee involvement. Restricted knee ROM, however, may reflect a number of factors, some related to the state of the articular cartilage, others to the joint capsule or adjacent muscle and ligamentous structures. During flexion the relative relationship between the axes of the femur and tibia change such that the traditional mechanical long arm goniometer is not an ideal instrument. Nevertheless, reliable determinations of knee ROM can be made by appropriately trained assessors[8-10]. In general, the measure has detected statistically significant differences between active treatments and either placebos or NSAID-free washout periods[13]. Disparity may exist between the number of degrees of movement of the knee and its functional capacity; knowledge of knee ROM may therefore lack clinical relevance.

Intermalleolar straddle. A measure of hip abduction, the intermalleolar straddle is reliable when determined by trained assessors[8-10]. Its responsiveness has been proved in NSAID trials[13]. However, like other measures of joint geometry, its relevance in pharmacodynamic studies is questionable and its use not recommended.

Intercondylar distance. A multiplanar measure of composite hip (abduction/external rotation) and knee (flexion) movement, this measurement can be performed reliably by trained assessors[8-10]. It has been shown capable of detecting statistically significant differences in one NSAID trial reviewed[13]. The relevance of the intercondylar distance is questionable for pharmacodynamic studies and its use is not recommended.

Measurement of joint geometry by ROM techniques generally requires assessor training, and like other performance based measures (e.g., 50 foot walk time and ascent time) the functional consequence of measured improvements or deteriorations is variable and difficult to quantitate. However, as

Table 1. Candidate variables for OA outcome assessment.

Variables	Potential Applicability			
	Knee	Hip	Hand	Generalized
Clinical				
Pain scales				
VA	+	+	+	+
Likert	+	+	+	+
Stiffness				
Severity	+	+	+	+
Duration	+	+	+	+
Redness	–	–	?	–
Physical performance				
Grip strength	–	–	+	–
50' walk time	+	+	–	?
Ascent time	?	?	–	?
Knee ROM	+	–	–	–
Intermalleolar straddle	–	+	–	–
Intercondylar distance	–	+	–	–
Swelling				
Arthrocircametry	?	–	?	–
Graded	+	–	?	–
Health status instruments				
WOMAC	+	+	–	–
Lequesne	+	+	–	–
Dreiser, *et al* (hand)	–	–	+	–
Bellamy, *et al* (hand)	–	–	+	–
HAQ	+	+	+	+
AIMS	+	+	+	+
AIMS2	+	+	+	+
FSI	+	+	+	+
Health related quality of life				
SF-36	?	?	?	?
EuroQol	?	?	?	?
NHP	?	?	?	?
HUI	?	?	?	?
Imaging				
Plain radiographs				
Nonmicrofocal	+	+	+	?
Microfocal	+	+	+	?
MRI	?	?	?	–
Scintigraphy	?	?	?	?
Ultrasound	?	?	?	?
Arthroscopy	+	?	?	?
Biological markers				
Aggrecan (SFAGN)	?	–	–	–
846 epitope in aggrecan (SF846)	?	–	–	–
Stromelysin-1 (SFSLN)	?	–	–	–
Collagenase (SFCLN)	?	–	–	–
Tissue inhibitor of metalloproteinase (SFTIMP)	?	–	–	–
Procollagen II C-propeptide (SFPCIIC)	?	–	–	–
Serum keratin sulfate (SKS)	?	?	?	?

+: Appropriate, –: not appropriate, ?: uncertain applicability

measures of clinical intermediaries (with the possible exception of the intercondylar distance), they are satisfactory.

Doyle Index. Several articular indices have been developed for the assessment of patients with RA. In that disorder, issues of joint selection, grading of tenderness and swelling, weighting and aggregation procedures, and the use of reduced joint counts have been thoroughly addressed. By comparison, the Doyle Index (a modification of the Ritchie Index) has received scant attention, and alternative methods of performing joint counts in OA have not been explored. For example, signal joints, reduced joint counts, grading and weighting/aggregation issues, all deserve further consideration[14].

The most obvious application for an articular index is in the assessment of patients with generalized OA. It is not clear, however, whether it is necessary to select and grade

Table 2. Concentrations of molecular markers of cartilage matrix turnover assayed in joint fluid and serum, with calculations of specificity and sensitivity for discrimination between presence and absence of knee joint pathology. Joint pathology in this context includes diagnosed injury to cruciate ligament and/or meniscus in the presence or absence of osteoarthritic joint changes detected by arthroscopy or radiography. (Reprinted with permission from Lohmander S: *Acta Orthop Scand 1995;*(suppl 266)66:84–7.)

	Vol[a]	SFAGN[b]	SF846[c]	SFSLN[d]	SFCLN[e]	SFTIMP[f]	SFPCIIC[g]	SKS[h]
Joint pathology (n)	2352	2119	385	1037	614	1028	428	758
Median	5	66	0.6	21	0.6	15	3.4	293
10th percentile	0.5	30	0.4	2.7	0	5	0.9	196
90th percentile	50	204	1.0	137	8.1	55	10.1	426
Reference (n)	118	88	9	77	26	77	49	137
Median	1	70	0.3	4.7	0.1	5	1.7	277
10th percentile	0.2	32	0.2	0.4	0	1.9	0.8	196
90th percentile	1.9	102	0.4	23.4	0.4	12	6.1	422
Specificity %*	83	83	82	83	84	84	83	84
Sensitivity %*	75	59	91	69	76	78	59	57

[a] Total volume of joint fluid aspirated (ml).
[b] Synovial fluid aggrecan fragments detected by Alcian blue precipitation (μg/ml).
[c] Synovial fluid 846 epitope in aggrecan detected by immunoassay (μg/ml).
[d] Synovial fluid stromelysin-1 (MMP-3) protein detected by immunoassay (nM).
[e] Synovial fluid collagenase (MMP-1) protein detected by immunoassay (nM).
[f] Synovial fluid tissue inhibitor of metalloproteinase (TIMP-1) protein detected by immunoassay (nM).
[g] Synovial fluid procollagen II C-propeptide detected by immunoassay (ng/ml).
[h] Serum keratan sulfate (5D4-epitope) detected by immunoassay (ng/ml).
* Specificity and sensitivity are calculated as sensitivity = (a/(a+c)) and specificity = (d/(b+d)), where the arbitrary cutoff point was set equal to the 80th percentile of the values for the reference group and where

	Disease	
	Present	Absent
Test positive	a	b
Test negative	c	d

the degree of tenderness in involved joints, or whether a simple count of the number of tender joints would suffice. Indeed, the weighting system developed by Lansbury[15] may be more relevant to OA than RA patients, given that the areas used in that weighting system more clearly approximate the cartilage area than that of the synovial membrane. Such counts also may be applicable to studies of hand OA. However, the relevance of using the index in patients whose disease is limited to the hips or knees (particularly monoarticular OA) is questionable, especially if the joint is simply graded as being either tender or nontender.

The assessment of swelling in OA is contentious, since it may be of bony or soft tissue origin. Furthermore, achieving adequate reliability in grading swelling may prove problematic. Since no pharmacologic agents have been developed that reverse the bony changes, it may not be appropriate to grade bony swelling at present. However, evaluation of the Doyle Index, and various modifications, may be warranted since the assessment can be performed by trained assessors in a reliable fashion and the index is capable of detecting change in clinical trials[8–10]. Possibly the applicability of the electronic dolorimeter in patients with OA could be evaluated[16].

Redness. Although the assessment of redness is mentioned in the US Food and Drug Administration (FDA) guidelines for outcome measurement, redness is found to be infrequent except in erosive arthritis[8]. In the early stage of nodal OA, erosive OA, and certain forms of secondary OA, redness sometimes occurs. However, even in these patients the assessment of redness probably adds little to the assessment process.

Grip strength. Grip strength is most commonly assessed using a pneumatic dynamometer (i.e., modified sphygmomanometer cuff coupled to a standard manometer)[17]. This method has been applied infrequently in OA clinical trials, possibly because of the relative infrequency with which the OA hand has been the focus of study[13]. The determination can be performed reliably by trained assessors[8–10]. In these studies it is necessary to consider whether grip strength should be determined with or without inclusion of counterpressure from the thumb. First carpometacarpal joint

involvement may be more readily assessed with a pinch meter. Grip strength determination may be important in hand OA studies and some generalized OA studies. However, the functional consequence of grip strength decrements is variable and may reduce the value of this measurement.

Arthrocircametry. Although employed in several studies, the measurement of joint circumference in centimeters has not proven a popular technique for evaluative research of antirheumatic drugs[13]. For the measurement of small hand joints, jeweller's rings or a flexible plastic tape can be employed[17]. In OA, soft tissue changes and effusions might improve with treatment (cf. bony proliferation). Even in RA clinical trials, where experience with arthrocircametry is much greater, there is little current enthusiasm for its routine incorporation[8–10,17]. As a result, I would not recommend the use of arthrocircametry in OA clinical trials.

Analgesic counts. In those studies in which pain is an endpoint, a measurement of supplementary analgesic use is usually essential to account for the potential effects of this important co-intervention. Methods occasionally used to correct pain scores for analgesic consumption are generally unsatisfactory. In short term studies, drug diaries may be a suitable method. However, in longer term trials, direct analgesic counts of the quantity consumed may be more appropriate. The use of pill container systems employing a microchip in the lid (to record opening and closing of the container) requires further evaluation.

Global assessments. A variety of types of global assessments may be made by either patient or physician. The following physician global assessments have been employed: overall assessment of physical disability, investigator's subjective opinion of each general condition, physician estimate of disease activity, and physician's global assessment of disease activity[8–10,12]. In contrast, the following patient global assessments have been employed: patient's overall assessment of pain, patient's overall assessment of physical disability, patient's estimate of disease activity, patient's opinion of general condition, patient's global assessment of disease activity[8–10,13]. Given the precedent set by RA clinical trials guidelines, and the frequent use of global assessments in routine clinical practice, it is appropriate to retain global assessments as candidate variables for outcome assessment in OA clinical trials.

Health status instruments. Health status instruments can be divided into those that are disease specific and those that are generic. This is a somewhat artificial division, but it serves to differentiate instruments with a relatively specific application from those used to assess arthritis overall or arthritis affecting many different joints. That is not to say that all generic instruments measure the same aspects of health or that any is a comprehensive measure of patient quality of life. In general, financial well being, personal security, religious freedom, and nutritional adequacy are not assessed. However, the majority of generic instruments, to a variable extent, measure health related quality of life.

Disease specific measures. Several measures have now been developed specifically to assess the effect of OA in particular joints.

Western Ontario and McMaster Universities Osteoarthritis (WOMAC) Index. The WOMAC Index is a tridimensional, disease specific, self-administered, health status measure[18]. It probes clinically important, patient relevant symptoms in the areas of pain, stiffness, and physical function in patients with OA of the hip and/or knee. The index consists of 24 questions (5 pain, 2 stiffness, 17 physical function) and can be completed in less than 5 minutes. It is available in both Likert (WOMAC LK3.0) and VAS (WOMAC VA3.0) formats. WOMAC is a valid, reliable, and responsive measure that can detect clinically important changes in health status after a variety of interventions (pharmacologic, surgical, physiotherapy, etc.). It has been translated into most European languages and has been requested for use by more than 130 researchers in 14 different countries. It has been recommended recently as a measure for assessing future slow acting drugs in OA (SADOA) clinical trials[19]. The following properties of the WOMAC instrument have been examined: Likert versus VAS, prior score availability, time frame dependency, single versus aggregate measurement, parametric versus nonparametric analysis, relative efficiency, weighting and aggregation, computerization of WOMAC VA3.0, and back translation and validation of foreign language translations. A user's guide is available for the WOMAC Index[20].

Lequesne algofunctional indices. Two OA algofunctional indices have been developed by Lequesne, one applicable to the hip, the other to the knee[21,22]. The indices contain 3 components: pain or discomfort, maximum distance walked, and activities of daily living. (A sexual function question included in the hip index is not considered necessary for antirheumatic drug studies.) The 2 indices are identical with respect to 4 of the 5 pain items and distance walked, but differ in the sitting pain and activities of daily living items. Points are allocated according to response such that higher values indicate greater severity. The indices are recommended as measures for OA trials in the 1985 European League Against Rheumatism (EULAR) Guidelines for antirheumatic drug research[23] and in the recent SADOA guidelines[19]. The reliability, validity, and responsiveness of the Lequesne algofunctional indices have been established. Although conceptually different from the WOMAC Index, the relative statistical efficiency of the WOMAC and Lequesne indices is very similar[20]. The following clinimetric issues of the Lequesne Index require some clarification: (1) how the question inventory was selected; (2) whether the index can be self-administered or requires an interviewer;

(3) what determines the score assigned for the different degrees of difficulty lying between the extremes of 0 and 2; (4) what time frame is used; (5) what conceptual and mathematical principles are used to aggregate pain, maximum distance walked, and activities of daily living into a single number; (6) can the break points for maximum distance walked be made mutually exclusive; (7) can item (a) in the pain component be modified to remove an ambiguity that relates to situations in which pain occurs without movement but also in certain positions. A generally available user's guide would be useful.

Dreiser algofunctional index. This index applicable to OA hand studies contains 10 items; 9 probe function and one explores the extent that a patient may be reluctant to accept a handshake[24]. This last question may be considered a pain related rather than a function related question. This is a relatively new index, and there has not been broad experience in its use. However, like the Lequesne Algofunctional Index, it represents an important contribution to clinical measurement. The Dreiser Index is physician administered, responses to each of the 10 questions being rated on 4 point verbal scales. Internal and external consistency, sensitivity and specificity, intraobserver reproducibility, responsiveness in placebo controlled trials, and ease of use have been assessed[24].

A novel tridimensional OA hand index. Our own research group is currently developing a tridimensional self-administered questionnaire probing pain, stiffness, and physical function in OA hand patients[25]. The development strategy has followed very closely that used for the WOMAC Index. The item inventory was generated from a combination of closed ended questions (derived from an examination of other indices and interviews with orthopedic surgeons, physiotherapists, and rheumatologists) and open ended questions. Responses to both types of questions were evaluated by interviewing 50 patients with hand OA. From that study, items of high prevalence, high frequency of recurrence, and moderate to extreme importance to patients were selected for incorporation into a test instrument. Both Likert (5 point) and VAS (10 cm horizontal) forms of the instrument have been prepared. Data on the reliability and construct validity of both formats have been obtained on 50 patients with hand OA. The responsiveness of the scale has yet to be determined. It is anticipated that the final index will be available in Summer 1997. Like the WOMAC Index, the new hand index can be self-administered and will be available in both Likert and VAS formats. Although index data will usually be analyzed on a scale by scale basis, we have used a device termed the PARIS Sectogram to derive weights for investigators who wish to aggregate the 3 subcomponents into a single score.

Health Assessment Questionnaire. Developed by Fries and co-workers at Stanford University, the Health Assessment Questionnaire (HAQ) has been well validated and is used widely[26]. The HAQ is self-administered, without any additional instructions, and the 2 dimensions of pain and disability can be completed in 5–8 minutes. The disability dimension has 8 categories: dressing and grooming, arising, eating, walking, hygiene, reach, grip, and activities (each containing 2 or 3 items). The instrument is valid, reliable, and responsive in a variety of different conditions and situations. It is particularly suitable for the assessment of generalized OA. Our own experience statistically comparing the WOMAC (pain and functional subscales) and the HAQ instruments in OA knee patients, is that, overall, the WOMAC may offer a slight advantage[27]. For this reason we have tended to use the WOMAC for OA hip and knee studies, and considered the HAQ useful in the assessment of generalized OA. The HAQ has been translated into several different languages. A user's guide is available. One of the advantages of the HAQ instrument is its brevity and its simplicity.

Arthritis Impact Measurement Scale (AIMS). The AIMS, developed by Meenan and co-workers in Boston, has, like the HAQ, been extensively validated[28]. It is a multidimensional, self-administered instrument using 46 items to probe 9 separate dimensions of mobility, physical activity, dexterity, social role, social activity, activities of daily living, pain, depression, and anxiety. The instrument is valid, reliable, and responsive and has been used in a variety of clinical settings. Our own experience statistically comparing the WOMAC Index and the AIMS has been that the WOMAC (pain and functional subscales) may overall offer slight advantage[27]. For this reason we have tended to employ the WOMAC in OA hip and knee studies, the AIMS, like the HAQ, being an excellent choice for studies of generalized OA. Meenan, *et al* have recently revised and expanded the AIMS instrument, producing a new questionnaire, AIMS2[29]. Nine different types of modification have been made to the original AIMS instrument. Reliability and validity determinations have been performed. Although a relatively long questionnaire, AIMS2 is a sophisticated instrument whose strength lies in its comprehensiveness and its self-administration. The new user's guide will be invaluable to those administering and scoring the AIMS and AIMS2 instruments. The only limitation of the AIMS instrument is the time for completion and the time requirements for scoring and analysis. It may not be necessary to use all subscales in all studies.

Functional Status Index (FSI). The FSI was developed by Jette and Deniston as part of the Pilot Geriatric Arthritis Project[30]. It measures the degree of dependence, pain, and difficulty experienced in performing daily activities. There are 2 forms of the FSI. The original contains 45 items spread across 3 dimensions, and takes 60–90 min to administer. The revised version contains 18 items grouped in 5 dimensions (gross mobility, hand activities, personal care, home chores and interpersonal activity). The shortened version

takes 20–30 min to administer. The instrument is interviewer administered (cf. HAQ, AIMS, WOMAC). The validity, reliability, and responsiveness of the instrument have been established, although it has not commonly been used in pharmacodynamic studies. Much of the early validation work was performed on RA rather than OA subjects according to a recent review of the instrument. The FSI may be applicable in assessing patients with generalized OA, although generally the HAQ or AIMS instruments have been more frequently employed.

Health related quality of life measures. A variety of instruments have been developed to probe various aspects of health related quality of life. Four instruments that may be applicable to OA clinical trials are: Short-Form (SF) 36 Health Status Questionnaire[31], EuroQol[32], the Health Utilities Index[33], and the Nottingham Health Profile[34].

SF-36 Health status questionnaire. This questionnaire, developed from the Rand Corporation's insurance experiment, is in the form of a self-administered questionnaire containing 36 items[31]. It takes about 5 min to complete and measures 3 major health attributes (functional status, well being, overall health) and 8 health concepts: (1) limitations in physical activities because of health problems; (2) limitations in social activities because of physical or emotional problems; (3) limitations in usual role activities because of physical health problems; (4) bodily pain; (5) general mental health; (6) limitations in usual role activities because of emotional problems; (7) vitality; and (8) general health perceptions. The questionnaire has been constructed for administration by telephone or by a trained interviewer. In a recent evaluation of patients undergoing total joint arthroplasty, Bombardier, *et al* showed that the WOMAC OA Index and the SF-36 provide complementary data[35], the former being a superior disease-specific instrument and the latter a superior generic instrument. Used together, these 2 high performance indices may be valuable.

European Quality of Life (EuroQol) Questionnaire. EuroQol is a self-administered questionnaire that classifies the patient into one of 243 health states[32]. It consists of a 5 part questionnaire probing deficits in mobility, self-care, main working activity, social relationships, pain and mood, and a VAS on which patients rate their own health status. It is suitable for use as a postal questionnaire, and does not require interviewer administration. It is recommended for use with other more detailed generic measures, such as the SF-36. There is relatively little experience with the EuroQol in pharmacodynamic evaluations of antirheumatic drugs.

Health Utilities Index. The Health Utilities Index was developed to provide a comprehensive description of health status in cancer patients[33]. The system measures 8 attributes: vision, hearing, speech, physical mobility, dexterity, cognition, pain and discomfort, and emotion. The index is self-administered and provides a single, overall summary score. It has been administered by face to face interview and also by telephone. There has been limited experience with the Health Utilities Index in pharmacodynamic evaluations of antirheumatic drugs in OA.

Nottingham Health Profile. The Nottingham Health Profile scale contains 38 items that can be grouped into 6 sections: physical mobility, pain, sleep, social isolation, emotional reactions, and energy level[34]. It is self-administered. Section scores may be presented as a profile or an overall score calculated. Experience with the Nottingham Health Profile in pharmacodynamic evaluations of antirheumatic drugs in OA is limited at the present time.

Experience with the measurement of health related quality of life in antirheumatic drug studies, particularly in OA, is extremely limited. It is, therefore, difficult, at the present time, to recommend one instrument over another. The experience of Bombardier, *et al*[35] suggests that the SF-36, when used in combination with the WOMAC Index, provides complementary data, and that this may, therefore, be the preferred combination for knee, and possibly for hip studies. No information is available on patients with hand OA, and we are not aware of evaluations performed specifically in patients with generalized OA. We require comparative studies of generic instruments in OA clinical trials and an examination of the extent to which other factors, such as social and emotional well being, helplessness, and comorbidities, might modulate the relationship between disease-specific instruments and health related quality of life instruments. At present my preference is to use the SF-36 until comparative data become available.

IMAGING TECHNIQUES

A "chondroprotective drug" has been defined as an agent that can retard, arrest, or reverse the degenerative process of OA in human hyaline cartilage[19]. To date, no pharmacologic agents have met this criterion. There has been debate as to how this phenomenon can be identified, and whether plain radiographs are adequate or whether alternative techniques, such as arthroscopy or magnetic resonance imaging (MRI), might be required. Indeed, the term "disease modifying" might be preferable to "chondroprotective," since pathologic changes also occur in the subchondral bone, synovial membrane, and adjacent tissues.

Plain radiographs. The literature of the radiology of OA joints has expanded rapidly in recent years[35–54]. Various methods have been described for taking the radiographs and a number of different scoring systems have been developed. Although somewhat lacking in sophistication, the Kellgren and Lawrence grading system can detect deterioration, albeit over a long period of time[37,48]. Newer, non-microfocal techniques, described by Dougados[36] and Buckland-Wright[43], based on specific views of the knee, can identify change over time. The advantage of the Dougados method is its simplicity, while the Buckland-Wright method relies on a

highly standardized procedure that requires radiographer training. The microfocal methods described by Buckland-Wright add a further level of sophistication, but may not be generally available for multicenter studies[45-47]. The microfocal techniques offer the advantage of high spatial resolution and high magnification, and have been used to explore osteoarthritic involvement of the hand and knee.

It is somewhat disconcerting that plain radiographs of the knee may not accurately predict the state of the articular cartilage as determined by arthroscopy[41,42]. Nevertheless, for purposes of randomized controlled trials, the aforementioned techniques are capable of detecting progressive joint space narrowing. Different methods have been described for measuring changes in joint space width or categorizing various degrees of narrowing.

MRI. The widespread use of MRI, for evaluative studies in OA, is constrained by cost and availability. Furthermore, there is only moderate concordance between articular cartilage lesions seen with MRI and those observed by arthroscopy[55]. Pilch, *et al* have identified errors in existing 3-D computer software used to perform volumetric studies of OA cartilage[56]. For clinical research purposes, MRI technology requires further development and validation.

Radionuclear scanning. Dieppe, *et al* have confirmed the value of scintigraphy in predicting subsequent loss of joint space in OA[57]. However, the role of scintigraphy in outcome assessment in OA clinical trials remains in doubt. Methods of quantitating the scintiscan have not been sufficiently refined, and the technique only examines subchondral bone and not the state of the adjacent cartilage. It seems likely, therefore, that radionuclide scanning will not be used for outcome assessment other than in predicting future radiographic change.

Ultrasonography. Myers, *et al* have shown *in vitro* that high frequency ultrasonic imaging provides accurate and reproducible measurements of the thickness and subsurface characteristics of human cartilage[58]. However, the capacity of this technology to detect chondroprotective effects has not been proved and remains in doubt. Validity and responsiveness, within the context of antirheumatic clinical trials in OA, have yet to be determined.

Arthroscopy. Although an invasive procedure, consideration has been given recently to the applicability of needle arthroscopic techniques in the evaluation of osteoarthritic cartilage in clinical trials. Ayral and co-workers have developed a new method for scoring chondroscopy and assessed its validity and reliability[59]. At the present time, however, the applicability of this technique, particularly in multicenter trials of potential chondromodulating agents in OA, has yet to be determined.

BIOLOGIC MARKERS

The destruction of joint cartilage in OA involves the degradation of matrix modules, which are then released into joint fluid, blood, and urine, where they may be detected by biochemical or immunological assays[60]. Such biologic markers might be of use in outcome measurement but the validity, reliability, and responsiveness of these markers have yet to be determined. Lohmander recently published an analysis of the sensitivity and specificity of biologic markers by comparing patients with joint pathology in the presence or absence of OA changes as determined by arthroscopy or radiography (Table 2). In general, the specificities are high, and the sensitivities moderate to high. When used in combinations of 2 or more markers (in the same sample), specificity and sensitivity may improve. For example, the combination of assays of stromelysin, collagenase, and tissue inhibitor of metalloproteinase in joint fluid results in a specificity of 93% and a sensitivity of 90%. At present there is insufficient relevant information on either synovial fluid (SF) or serum to be able to recommend one measure or combination of measures over alternatives. Indeed, it is not certain at this stage whether any of the aforementioned markers will prove suitable to assess change over time. Further work is required in this important area of molecular biology.

FUTURE PERSPECTIVE

Future decisions regarding outcome measurement should be based partly on prior experience and partly on recent developments. Few formal preferences have been expressed to date regarding clinical outcome measures for OA trials other than those encompassed in the FDA Guidelines (1988)[61], the EULAR[23], and the SADOA Guidelines[19] (Table 3). The FDA guidelines are not as explicit about outcome measurement in OA clinical trials as they are regarding RA and AS trials. The Guidelines do not clearly specify "primary efficacy variables," but suggest that "efficacy evaluation of other manifestations of disease, even though they may not be shared by all patients, should be carried out with respect to change in swelling, redness, tenderness on pressure, pain at rest or on motion, change in range of motion, and walking or stair climbing. In addition, the investigator's and patient's opinion of the patient's condition on the day of assessment are recommended." In contrast, the EULAR guidelines recommend the following efficacy measures: index of severity of hip and knee disease (Lequesne algofunctional indices), investigator's overall opinion, pain VAS, patient's overall opinion, and walking time (if knee, stair climbing is recommended). The SADOA Guidelines recommend the use of a pain scale, functional indices (Lequesne or WOMAC), Doyle Index, loss of mobility, walking or stair climbing time, global assessment of efficacy by patient, global assessment of efficacy by physician, number of flares, analgesic consumption, simplified quality of life instrument.

A review of prior NSAID clinical trials in OA indicates a preference for measures of pain, patient global assessment, and physician global assessment. However, functional dis-

Table 3. Guidelines for OA clinical trials.

EULAR[23]

1. Index of severity of hip and knee disease (Lequesne)
2. Investigator's overall opinion
3. Pain on VAS
4. Patient's overall opinion
5. Walking time (if knee, stair climb recommended)

FDA[54]

1. Swelling
2. Redness
3. Tenderness on pressure
4. Pain at rest or on motion
5. Range of motion
6. Walking or stair climbing time
7. Investigator's opinion of the patient's condition on the day of assessment
8. Patient's opinion of his/her condition on the day of assessment

SADOA[19]

1. VAS
2. Functional indices (WOMAC or Lequesne)
3. Doyle Index
4. Loss of mobility
5. Walking or stair climbing time
6. Consumption of analgesics and/or NSAID
7. Number of flares over time, especially effusion in OA knee
8. Patient's overall judgment of efficacy
9. Investigator's overall judgment of efficacy
10. Quality of life scale

(Reprinted with permission from: Bellamy N: Instruments to assess osteoarthritis — current status and future needs (editorial). *Ann Rheum Dis 1995;*54:692–3.)

ability, the 2nd most important consequence of OA, which is amenable to measurement and has a clinical importance and patient relevance greater than that of measures of joint geometry or physical performance tests, has been assessed with relative infrequency. My current preference for clinical outcome measures is illustrated in Table 4. In particular, I favor measures of pain, function, patient global assessment, physician global assessment, and a generic health status measure. Comparative studies of generic health status measures will be required before a preferred measure can be identified, but at present the SF-36 appears to be suitable. There are currently no international guidelines for imaging procedures or biologic marker assays for OA clinical trials. With respect to radiographic techniques, standardized methods, such as those described by Buckland-Wright (nonmicrofocal) or Dougados are of immediate application and do not impose the level of operational constraints required for microfocal methods. Joint space narrowing can be accurately assessed by a perspex ruler, or a grid, or by computer assisted methods. Alternatively, the grading systems, based on photographic standards, recently published by Altman, *et al*[62] or Burnett, *et al*[63] are superior to the original classification system advanced by Kellgren and Lawrence[48]. While MRI technology may provide future opportunities, further validation work is required, particularly with respect to volumetric analysis. Needle arthroscopy may have a role to play, but its invasive nature and its limited availability make it currently unsuitable for widespread application in multicenter studies[64]. The ideal biologic marker has yet to be identified, and it will require further developments in the area of molecular biology before consensus can be reached on standard measures of demonstrated reliability, validity, and responsiveness[65]. A serologic marker would be much easier to use than a SF marker. Although this is a difficult area of clinical metrology, progress in recent years has been extremely encouraging. International agreement on clinical

Table 4. Clinical outcome measurement batteries for OA trials.

Measure	Generalized OA	Hip OA	Knee OA	Hand OA
Pain	HAQ or AIMS, AIMS2	WOMAC	WOMAC	Bellamy, *et al**
Stiffness	VA or Likert	WOMAC	WOMAC	Bellamy, *et al**
Physical functional (indices)	HAQ or AIMS, AIMS2	WOMAC	WOMAC	Bellamy, *et al**
Global assessment	MD/patient	MD/patient	MD/patient	MD/patient
General health status/QOL	SF-36	SF-36	SF-36	SF-36

* The validation of this novel tridimensional index is nearing completion.

outcome measures can be achieved this year. Development in imaging techniques and molecular biology is progressively advancing.

REFERENCES

1. Bellamy N: Future perspective. In: *Musculoskeletal Clinical Metrology.* Dordrecht: Kluwer Academic Publishers, 1993:253–68.
2. Bellamy N: Pain measurement. In: *Musculoskeletal Clinical Metrology.* Dordrecht: Kluwer Academic Publishers, 1993:65–76.
3. Hart LE: Measuring the pain of arthritis — Validation of a new scale. MSc. Thesis, McMaster University, Hamilton, Ontario, 1993.
4. Grossi E, Borghi C, Cerchiari EL, *et al:* Analogue chromatic continuous scale (ACCS): A new method for pain assessment. *Clin Exp Rheumatol 1983;*1:337–40.
5. Melzack R: The McGill Pain Questionnaire: Major properties and scoring methods. *Pain 1975;*1:275–99.
6. Anderson KO, Bradley LA, McDaniel LK, *et al:* The assessment of pain in rheumatoid arthritis: Disease differentiation and temporal stability of a behavioral observation method. *J Rheumatol 1987;*14:700–4.
7. Anderson KO, Bradley LA, McDaniel LK, *et al:* The assessment of pain in rheumatoid arthritis — Validity of a behavioral observation method. *Arthritis Rheum 1987;*30:36–43.
8. Bellamy N: Standardised procedures for outcome measurement and parameters for calculating sample size for antirheumatic drug studies. In: *Musculoskeletal Clinical Metrology.* Dordrecht: Kluwer Academic, 1993:193–251.
9. Bellamy N: Standard deviation tables for sample size calculation in rheumatoid arthritis, osteoarthritis, ankylosing spondylitis and fibromyalgia clinical trials. In: *Musculoskeletal Clinical Metrology.* Dordrecht: Kluwer Academic Publishers, 1993:337–48.
10. Bellamy N: Estimated sample size requirements for clinical trials in rheumatoid arthritis, osteoarthritis, ankylosing spondylitis and fibromyalgia. In: *Musculoskeletal Clinical Metrology.* Dordrecht: Kluwer Academic Publishers, 1993:349–57.
11. Bellamy N, Buchanan WW: A preliminary evaluation of the dimensionality and clinical importance of pain and disability in osteoarthritis of the hip and knee. *Clin Rheumatol 1986;*5:231–41.
12. Bellamy N, Wells GA, Campbell J: PARIS Sectogram: A method for weighting and aggregating the WOMAC Osteoarthritis Index (abstr). *Osteoarthritis Cart 1994;*(suppl 1)2:37.
13. Bellamy N, Buchanan WW: Outcome measurement in osteoarthritis clinical trials: The case for standardisation. *Clin Rheumatol 1984;*3:293–303.
14. Doyle DV, Dieppe PA, Scott J, *et al:* An articular index for the assessment of osteoarthritis. *Ann Rheum Dis 1981;*40:75–8.
15. Bellamy N: Articular indices. In: *Musculoskeletal Clinical Metrology.* Dordrecht: Kluwer Academic Publishers, 1993:103–16.
16. Atkins CJ, Zielinski A, Klinkhoff AA, *et al:* An electronic method for measuring joint tenderness in rheumatoid arthritis. *Arthritis Rheum 1992;*35:407–10.
17. Bellamy N: Mechanical and electromechanical devices. In: *Musculoskeletal Clinical Metrology.* Dordrecht: Kluwer Academic Publishers, 1993:117–34.
18. Bellamy N, Buchanan WW, Goldsmith CH, *et al:* Validation study of WOMAC: A health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol 1988;*15:1833–40.
19. Lequesne M, Brandt K, Bellamy N, *et al:* Guidelines for testing slow acting drugs in osteoarthritis. *J Rheumatol 1994;*(suppl 41) 21:65–73.
20. Bellamy N: WOMAC Osteoarthritis Index — A User's Guide. London: London Health Sciences Centre, 1995.
21. Lequesne MG, Mery C, Samson M, *et al:* Indexes of severity for osteoarthritis of the hip and knee: Validation — Value in comparison with other assessment tests. *Scand J Rheumatol 1987;*(suppl 65):85–9.
22. Bellamy N: Health status instruments and functional indices. In: *Musculoskeletal Clinical Metrology.* Dordrecht: Kluwer Academic Publishers, 1993:77–101.
23. Guidelines for the clinical investigation of drugs used in rheumatic diseases, European drug guidelines, Series 5. World Health Organization, Regional Office for Europe. Copenhagen: European League Against Rheumatism. March 1985:7–24.
24. Dreiser R-L, Maheu E, Guillou JB, *et al:* Validation of an algofunctional index for osteoarthritis of the hand. *Rev Rhumatisme English ed. 1995;*(suppl 1)6:43S–53S.
25. Bellamy N, Klestov A, Muirden K, *et al:* Osteoarthritis in twins: Analysis of observer agreement for the clinical diagnosis of osteoarthritis (abstr). 26th Scandinavian Congress of Rheumatology. May 31–June 3, 1996, Reykjavik, Iceland.
26. Fries JF, Spitz P, Kraines RG, *et al:* Measurement of patient outcome in arthritis. *Arthritis Rheum 1980;*23:137–45.
27. Griffiths G, Bellamy N, Bailey WH, *et al:* A comparative study of the relative efficiency of the WOMAC, AIMS, and HAQ instruments in evaluating the outcome of total knee arthroplasty. *Inflammopharmacology 1995;*3:1–6.
28. Meenan RF, Gertman PM, Mason JH: Measuring health status in arthritis: The Arthritis Impact Measurement Scales. *Arthritis Rheum 1980;*23:146–52.
29. Meenan RF, Mason JH, Anderson JJ, *et al:* The content and properties of a revised and expanded Arthritis Impact Measurement Scales Health Status Questionnaire. *Arthritis Rheum 1992;*35:1–10.
30. Jette AM, Deniston OL: Inter observer reliability of a functional status assessment instrument. *J Chron Dis 1978;*31:573–80.
31. Ware JE Jr, Sherbourne CD: The MOS 36-item Short-Form Health Survey (SF-36). I. Conceptual frame-work and item selection. *Med Care 1992;*30:473–81.
32. Hurst NP, Jobanputra P, Hunter M, and the Economic and Health Outcomes Research Group: Validity of EuroQol — A generic health status instrument — in patients with rheumatoid arthritis. *Br J Rheumatol 1994;*33:655–62.
33. Feeny D, Furlong W, Barr RD, *et al:* A comprehensive multiattribute system for classifying the health status of survivors of childhood cancer. *J Clin Oncol 1992;*10:923–8.
34. Hunt SM, McKenna SP, McEwen J, *et al:* The Nottingham Health Profile: Subjective health status and medical consultations. *Soc Sci Med 1981;*15A:221–9.
35. Bombardier C, Melfi CA, Paul J, *et al:* Comparison of a generic and a disease-specific measure of pain and physical function after knee replacement surgery. *Med Care 1995;*(suppl)33:AS131–AS144.
36. Dougados M, Guegnen A, Nguyen M, *et al:* Longitudinal radiologic evaluation of osteoarthritis of the knee. *J Rheumatol 1992;*19:378–84.
37. Spector TD, Dacre JE, Harris PA, Huskisson EC: Radiological progression of osteoarthritis: An 11 year follow up study of the knee. *Ann Rheum Dis 1992;*51:1107–10.
38. Cooper C, Cushnaghan J, Kirwan JR, *et al:* Radiographic assessment of the knee joint in osteoarthritis. *Ann Rheum Dis 1992;*51:80–2.
39. Jones AC, Ledingham J, McAliadon T, *et al:* Radiographic assessment of patellofemoral osteoarthritis. *Ann Rheum Dis 1993;*52:655–8.
40. Lane NE, Nevitt MC, Genant HK, Hochberg MC: Reliability of new indices of radiographic osteoarthritis of the hand and hip and lumbar disc degeneration. *J Rheumatol 1993;*20:1911–8.
41. Fife RS, Brandt KD, Braunstein EM, *et al:* Relationship between arthroscopic evidence of cartilage damage and radiographic

evidence of joint space narrowing in early osteoarthritis of the knee. *Arthritis Rheum 1991;*34:377–82.
42. Blackburn WD Jr, Bernreuter WK, Rominger M, Loose LL: Arthroscopic evaluation of knee articular cartilage: A comparison with plain radiographs and magnetic resonance imaging. *J Rheumatol 1994;*21:675–9.
43. Buckland-Wright JC: Quantitative radiography of osteoarthritis. *Ann Rheum Dis 1994;*53:268–75.
44. Theiler R, Ghosh P, Brooks P: Clinical, biochemical and imaging methods of assessing osteoarthritis and clinical trials with agents claiming "chondromodulating" activity. *Osteoarthritis Cart 1994;*2:1–23.
45. Buckland-Wright JC, MacFarlane DG, Lynch JA: Relationship between joint space width and subchondral sclerosis in the osteoarthritic hand: A quantitative microfocal radiographic study. *J Rheumatol 1992;*19:788–95.
46. Buckland-Wright JC, MacFarlane DG, Lynch J: *Quantitative* microfocal radiographic assessment of disease and progression in osteoarthritis of the hand. *J Rheumatol 1991;*(suppl 27)18:40–1.
47. Buckland-Wright JC, MacFarlane DG, Jasani MK, Lynch JA: Quantitative microfocal radiographic assessment of osteoarthritis of the knee from weight bearing tunnel and semiflexed standing views. *J Rheumatol 1994;*21:1734–41.
48. Kellgren JH: The epidemiology of chronic rheumatism. In: *Atlas of Standard Radiographs,* vol. 2. Oxford: Blackwell Scientific, 1963.
49. Buckland-Wright JC, MacFarlane DG, Williams SA, Ward RJ: Accuracy and precision of joint space width measurements in standard and macroradiographs of osteoarthritic knees. *Ann Rheum Dis 1995;*54:872–80.
50. Buckland-Wright JC: Protocols for precise radio-anatomical positioning of the tibio-femoral and patello femoral compartments of the knee. *Osteoarthritis Cartilage 1995;*(suppl A)3:71–80.
51. Buckland-Wright JC: Quantitation of radiographic changes. In: Brandt KD, Lohmander S, Doherty M, eds. *Osteoarthritis.* Oxford: Oxford University Press, 1997 (in press).
52. Buckland-Wright JC, MacFarlane DG, Lynch JA, Jasani MK: Quantitative microfocal radiography detects changes in OA knee joint space width in patients in placebo-controlled trial of NSAID therapy. *J Rheumatol 1995;*22:937–43.
53. Dieppe P, Cushnaghan J, McAlindon T: Epidemiology, clinical course and outcome of knee osteoarthritis. In: Kuettner K, Schleyerbach R, Peyron JG, Hascall VC, eds. *Articular Cartilage and Osteoarthritis.* New York: Raven Press, 1992:617–27.
54. Buckland-Wright JC, MacFarlane DG, Lynch JA, Jasani MK, Bradshaw CR: Joint space width measures cartilage thickness in osteoarthritis of the knee: High resolution plain film and double contrast microradiographic investigation. *Ann Rheum Dis 1995;*54:263–8.
55. Adams ME, Li DKB, McConkey JP, *et al:* Evaluation of cartilage lesions by magnetic resonance imaging at 0.15 T: Comparison with anatomy and concordance with arthroscopy. *J Rheumatol 1991;*18:1573–80.
56. Pilch L, Stewart C, Gordon D, *et al:* Assessment of cartilage volume in the femorotibial joint with magnetic resonance imaging and 3D computer reconstruction. *J Rheumatol 1994;*21:2307–19.
57. Dieppe P, Cushnaghan J, Young P, Kirwan J: Prediction of the progression of joint space narrowing in osteoarthritis of the knee by bone scintigraphy. *Ann Rheum Dis 1993;*52:557–63.
58. Myers SL, Dines K, Brandt DA, Brandt KD, Albrecht ME: Experimental assessment by high frequency ultrasound of articular cartilage thickness and osteoarthritic changes. *J Rheumatol 1995;*22:109–16.
59. Ayral X, Dougados M, Listrat V, *et al:* Chondroscopy: A new method for scoring chondropathy. *Semin Arthritis Rheum 1993;*22:289–97.
60. Lohmander LS, Roos H, Dahlberg L, *et al:* The role of molecular markers to monitor disease intervention and cartilage breakdown in osteoarthritis. *Acta Orthop Scand 1995;*(suppl 266)66:84–7.
61. Guidelines for the clinical evaluation of anti-inflammatory and antirheumatic drugs (adults and children). US Department of Health and Human Services; Public Health Service, Food and Drug Administration. Rockville: April 1988:12–5.
62. Altman RD, Hochberg M, Murphy WA Jr, Wolfe F, Lequesne M: Atlas of individual radiographic features in osteoarthritis. *Osteoarthritis Cart 1995;*(suppl A)3:3–70.
63. Burnett S, Hart DJ, Cooper C, Spector TD: *A Radiographic Atlas of Osteoarthritis.* London: Springer-Verlag, 1994.
64. Blackburn WD Jr, Chivers S, Bernreuter W: Cartilage imaging in osteoarthritis. *Semin Arthritis Rheum 1996;*25:273–81.
65. Westacott CI, Sharif M: Cytokines in osteoarthritis: Mediators or markers of joint destruction. *Semin Arthritis Rheum 1996;*25:254–72.

Recommendations for a Core Set of Outcome Measures for Future Phase III Clinical Trials in Knee, Hip, and Hand Osteoarthritis. Consensus Development at OMERACT III

NICHOLAS BELLAMY, JOHN KIRWAN, MAARTEN BOERS, PETER BROOKS, VIBEKE STRAND, PETER TUGWELL, ROY ALTMAN, KENNETH BRANDT, MAXIME DOUGADOS, and MICHEL LEQUESNE

ABSTRACT. Significant progress has been made in outcome measurement procedures for osteoarthritis (OA) clinical trials, and guidelines have been established by the US Food and Drug Administration, European League Against Rheumatism, the World Health Organization/International League of Associations for Rheumatology, and the Group for the Respect of Ethics and Excellence in Science. However, there remains a need for further international harmonization of measurement procedures used to establish beneficial effects in Phase III clinical trials. A key objective of the OMERACT III conference was to establish a core set of outcome measures for future phase III clinical trials. During the conference, using a combination of discussion and polling procedures, a consensus was reached by at least 90% of participants that the following 4 domains should be evaluated in future phase III trials of knee, hip, and hand OA: pain, physical function, patient global assessment, and, for studies of one year or longer, joint imaging (using standardized methods for taking and rating radiographs, or any demonstrably superior imaging technique). These evidence based preferences, achieved with a high degree of consensus, establish an international standard for future phase III trials and will also facilitate metaanalysis and Cochrane Collaborative Project goals. (*J Rheumatol 1997;24:799–802*)

Key Indexing Terms:
OSTEOARTHRITIS ENDPOINTS CORE SET OUTCOME MEASURES

Outcome measurement in clinical trials requires the use of valid, reliable, and responsive measurement procedures that adequately capture important aspects of the condition. In recognition of this requirement, a number of individuals and groups have published lists of recommended outcome measures[1–5]. In particular, the US Food and Drug Administration, European League Against Rheumatism, World Health Organization/International League of Associations for Rheumatology, and the Group for Respect of Ethics and Excellence in Science have published guidelines which in part specify domains and in part recognize actual measurement techniques or instruments. While not in complete agreement, the existing guidelines nevertheless share several important elements, namely, the measurement of pain, walk time, patient global assessment, and physician global assessment.

To build on experience and current preference but not exclude other measures of potential importance in future trials, a process was followed that had 4 basic elements: (1) provision of information from the literature; (2) lectures followed by discussion periods; (3) breakout groups; (4) polling procedures.

THE PROCESS

Prior to OMERACT III, participants were asked to complete an initial questionnaire to identify candidate variables. From return questionnaires, a second questionnaire was then constructed incorporating additional suggestions. The questionnaire was extensive and identified 4 site specific forms of osteoarthritis (OA) (knee, hip, hand, and generalized), 2 types of studies (symptom modifying OA drugs and structure modifying OA drugs), 3 levels of measurement (clinical, imaging, and biologic markers) and various domains and measurement techniques. Participants were asked to rank in order of importance their preferences for outcome measurement for each clinical situation and drug class. This proved excessively demanding and only 15 questionnaires were returned. Prior to OMERACT III each participant also received position papers that outlined the dimensionality of

From London Health Sciences Centre, Victoria Campus, London, Canada; Bristol Royal Infirmary, Bristol, UK; University Hospital Maastricht, Maastricht, The Netherlands; St. Vincent's Hospital, Darlinghurst, Australia; Stanford University, San Francisco, USA; University of Ottawa Ottawa, Canada; Miami School of Medicine, Miami, USA; Indiana University School of Medicine, Indianapolis, USA; Hôpital Cochin and College de Medecine des Hôpitaux de Paris, Paris, France.

N. Bellamy, MD, Victoria Hospital; J. Kirwan, MD, Royal Bristol Infirmary; M. Boers, University Hospital Maastricht; P. Brooks, MD, St. Vincent's Hospital; V. Strand, Stanford University; P. Tugwell, MD, University of Ottawa; R. Altman, MD, Miami School of Medicine; K. Brandt, Indiana University School of Medicine; M. Dougados, MD, Hôpital Cochin; M. Lequesne, MD, College de Medecine des Hôpitaux de Paris.

Address reprint requests to Dr. N. Bellamy, Ste. 303 Colborne, Bldg., London Health Sciences Centre, 375 South Street, London, ON N6A 4G5, Canada.

the measurement problem and provided up-to-date information in areas of clinical, imaging, and biologic markers.

During OMERACT III, participants attended presentations addressing the different measurement areas and, where available, data were presented on the clinimetric properties of different instruments and comparisons of measurement techniques. Time was allowed during question period for clarification and for alternative viewpoints. Participants then completed an exercise in which they were asked to assign 100 points to reflect their measurement preferences in each of 4 types of OA trials (knee, hip, hand, generalized). Participants next separated into 3 clinical and one combined imaging/biologic markers group. Feedback was available from the voting profiles within each of the breakout groups. The breakout groups provided an opportunity to discuss contentious issues more fully and bring back recommendations to the group as a whole. Following these deliberations as well as other informal discussions, a final questionnaire was designed to allow participants to vote for inclusion of domains in a core set and to express use preferences for types of instruments. However, questions regarding specific instruments, while permitting flexibility, were not generated from prior voting procedures and a decision was made not to include recommendations regarding specific instruments for research applications.

THE CONSENSUS

Participants were provided opportunity to recommend a measure for inclusion in (a) the core set (i.e., mandatory in future Phase III clinical trials in knee, hip, and hand OA studies); (b) the research agenda (i.e., worthy of further formal evaluation and possible future inclusion in the core set); or (c) inclusion in neither the core set nor the research agenda. The summary results are shown in Table 1.

After presentation of these data a number of issues were raised.

1. Whether generalized OA was a distinct and definable entity for clinical trials purposes. (Resolution — to exclude further consideration of generalized OA.)

2. Whether the rate of onset of therapeutic effect (fast versus slow) determined the need for different types of clinical measures. (Resolution — time of onset determines when to measure rather than what to measure.)

3. Whether different measures were required for an analgesic study versus a nonsteroidal antiinflammatory drug (NSAID) study. (Resolution — the domains are the same but the measurement techniques might vary.)

4. Whether clinical measures should be different for system modifying versus structure modifying OA drug studies. (Resolution — the clinical core domains are the same.)

5. It was assumed that biologic markers would be important in the future, but confirmatory evidence is lacking for the evaluative and predictive value of any single marker.

6. It was acknowledged that data existed on the value of measures of health related quality of life (generic and utility measures), but that no one measure had yet been identified as superior to all others for clinical trial purposes. The importance of such measures in health related quality of life determination, cross study and cross disease comparisons, and in pharmacoeconomic comparisons was generally acknowledged. As a result, while not in the core set, it was decided to strongly recommend the incorporation of health related quality of life measures in future Phase III trials of at least 6 months' duration. Over the next 3 to 5 years it should be possible to evaluate the role of such measures in clinical trials.

7. It was emphasized that no measure was excluded from use in future clinical trials by decisions made at OMERACT III. Indeed, in some studies the primary outcome might not be one cited in the core set (e.g., the effect of a future drug on time to surgery). However, such studies would be required to also include assessments of domains cited in the core set in the measurement battery.

8. There was debate whether stiffness should be incorporated, whether pain and stiffness were part of the same domain, whether patients understood the concept of stiff-

Table 1. Preferences for core set of efficacy domains in future Phase III hip, knee, and hand OA trials.

Domain	In Core (% Voting Yes)	In Research Agenda (% Voting Yes)	In Neither (% Voting Yes)	Number Voting
Pain	100	0	0	75
Physical function	97	1	1	76
Imaging* (in studies of 1 yr or longer)	92	7	1	76
Patient global assessment	91	1	1	75
Physician global assessment	52	21	27	73
Generic quality of life/utility	36	58	6	69
Stiffness	14	61	25	72
Other**	13	69	19	16
Inflammation	8	70	22	74

* Standardized techniques for taking and scoring radiographs or demonstrably superior imaging techniques.
** Includes tenderness, performance based measures, time to surgery, number of flares, biologic markers.

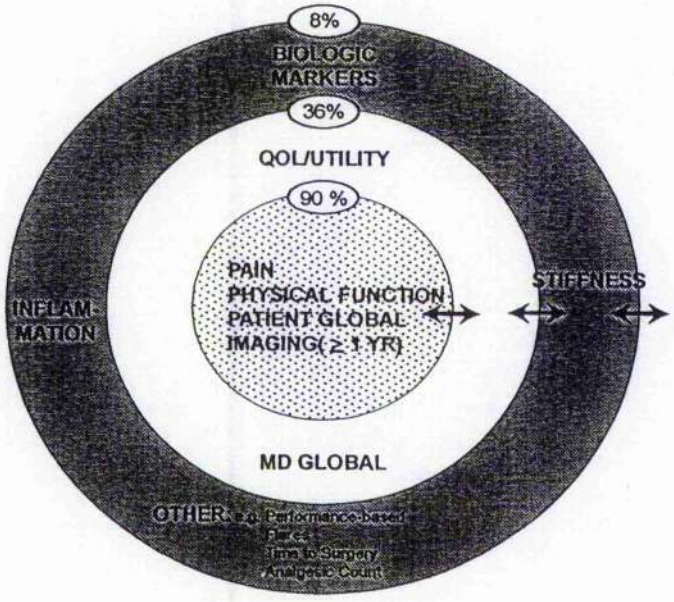
ness, and whether current techniques accurately assessed it. (Resolution — when stiffness is to be assessed in hip and/or knee studies it should be measured using the WOMAC or Algofunctional Severity Indices.)

9. There was debate on the value of physician global assessment in OA trials (as there had been at OMERACT I regarding its use in rheumatoid arthritis trials). Only 52% of participants felt it should be included in the core set for OA and as a result it was not included. It was acknowledged, however, that it was important to about half the participants and its continued use was acceptable.

In drawing up the core set, 3 assumptions were proposed: (1) to be included there needed to be evidence for reliability, validity, and responsiveness; (2) it was not necessary to specify exact instruments, but only to agree on the major domains to be included; (3) there is a difference between consensus and unanimity. However, a 51/49% split seemed insufficient, since 49% of participants would be in disagreement. Similarly, a 60/40% split would not be decisive. Common sense suggests if 90% or more participants agreed on a core set, one could claim a consensus, albeit without unanimity. As a result the core set recommended by OMERACT III was based on a consensus of ≥ 90% and included the following measures:

- Pain
- Physical function
- Patient global assessment
- Imaging in studies ≥ 1 year (As an efficacy measure in structure modifying OA drug studies, but also as a safety measure in pure system modifying OA drug studies of ≥ 1 year duration)

These are illustrated in Figure 1, in which the inner core defines the core set for OA. The middle core identifies health related quality of life measures (optional, but strongly recommended) and physician global assessment (optional, depending on perceived importance to the investigator). The outer core contains measures of stiffness (by WOMAC and Algofunctional Severity Indices), biologic markers, measures of inflammation, and other assessments (e.g., performance based measures, flares, time to surgery, analgesic consumption), all of which are optional measures. This concept places highly patient relevant measures at the center, while measures less relevant to patients are at the periphery. It should be noted that only domains cited in the inner core (i.e., core set) will be obligatory in outcome measurement in future Phase III trials. Any instrument used should be of adequate reliability, validity, and responsiveness. For imaging, the preferred technique currently is radiographic and



% voting for inclusion in core set	Placement	Consequence
≥ 90%	INNER CORE →	"CORE SET"
≥ 36% - < 90%	MIDDLE CORE →	QOL/UTILITY(Strongly Recommended)
8% - < 36%	OUTER CORE →	OPTIONAL

Figure 1. Osteoarthritis core concept.

requires standardized methods for both taking and scoring films. The term imaging was selected specifically to allow for future developments of technically superior methods.

CONCLUSION

These evidence based preferences were achieved through a high degree of consensus. They allow international harmonization of outcome measurement procedures in OA clinical trials. However, they also offer 4 additional advantages: (1) they do not exclude other measures being used in addition to the core set; (2) they are flexible and allow over time for the inward and outward migration of measures as developments occur in clinical, imaging, and molecular disciplines; (3) they create a foundation on which other organizations and consensus conferences can build, particularly with respect to the specification of exact instruments for use in specific situations; and (4) they will facilitate metaanalyses and Cochrane Collaborative Project goals[6].

In summary, participants at OMERACT III agreed (≥ 90%) on a core set of 4 domains for outcome measurement in future Phase III clinical trials of hip, knee, and hand OA. The 4 domains identified were pain, physical function, patient global assessment, and, for studies of at least one year, joint imaging.

REFERENCES

1. Guidelines for the Clinical Evaluation of Anti-inflammatory and Anti-rheumatic Drugs (Adults and Children). US Department of Health and Human Services. Public Health Services, Food and Drug Administration. Rockville, MD: April 1988:12–5.
2. Guidelines for the Clinical Investigations of Drugs Used in Rheumatic Diseases. European Drug Guidelines, Series 5. World Health Organization, Regional Office for Europe, Copenhagen, European League Against Rheumatism. March 1985:21–4.
3. Lequesne M, Brandt K, Bellamy N, *et al:* Guidelines for testing slow acting drugs in osteoarthritis. *J Rheumatol 1994;*(suppl 41)21:65–73.
4. Group for the Respect of Ethics and Excellence in Science (GREES): Osteoarthritis Section. Recommendations for the registration of drugs used in the treatment of osteoarthritis. *Ann Rheum Dis 1996;*55:552–7.
5. Bellamy N: Osteoarthritis. In: *Musculoskeletal Clinical Metrology.* Dordrecht: Kluwer Academic Publishers, 1993:169–76.
6. Cochrane AL: Effectiveness and Efficiency. Random Reflections on Health Services. London: Nuffield Provincial Hospitals Trust, 1972. (reprinted) London: *BMJ,* 1989.

Osteoarthritis and Cartilage (1996) 4, 217–243

1063-4584/96/040217+27 $12.00/0

OSTEOARTHRITIS and CARTILAGE

SPECIAL REPORT

DESIGN AND CONDUCT OF CLINICAL TRIALS IN PATIENTS WITH OSTEOARTHRITIS:

Recommendations from a task force of the Osteoarthritis Research Society

Results from a workshop

Steering Committee:
ROY ALTMAN, MIAMI, FL, U.S.A.—Chairman
KENNETH BRANDT, INDIANAPOLIS, IN, U.S.A.
MARC HOCHBERG, BALTIMORE, MD, U.S.A.
ROLAND MOSKOWITZ, CLEVELAND, OH, U.S.A.

Committee members:
NICHOLAS BELLAMY, CANADA
DANIEL A. BLOCH, STANFORD, CA, U.S.A.
JOSEPH BUCKWALTER, IOWA CITY, IO, U.S.A.
MAXIME DOUGADOS, PARIS, FRANCE
GEORGE EHRLICH, PHILADELPHIA, PA, U.S.A.
MICHEL LEQUESNE, PARIS, FRANCE
STEFAN LOHMANDER, LUND, SWEDEN
WILLIAM A. MURPHY, JR. HOUSTON, TX, U.S.A.
THERESA ROSARIO-JANSEN, CINCINNATI, OH, U.S.A.
BENJAMIN SCHWARTZ, ST. LOUIS, MO, U.S.A.
STEPHEN TRIPPEL, BOSTON, MA, U.S.A.

CONTENTS

Address correspondence to: Roy D. Altman, M.D. (D26), PO Box 016960, Miami, FL 33101, U.S.A. Overnight mail: Roy D. Altman, MD (NH207G), Miami Dept. Veterans Affairs Hosp., 1201 NW 16 St., Miami, FL 33125, U.S.A.

A. Introduction

There have been many recent advances in understanding the pathophysiology and evolution of osteoarthritis (OA). These advances have led to improvement in diagnosis and therapy, and have prompted a re-evaluation of the methodology and metrology involved in the performance of clinical trials in OA. Recently, a combined committee of the World Health Organization (WHO) and International League of Associations for Rheumatologists (ILAR) has defined two classes of symptomatic therapy based on the onset and duration of the response to treatment [1], and has proposed a third classification for agents that may alter the disease process. In addition, a workshop sponsored by the WHO and the American Academy of Orthopedic Surgeons (AAOS) has reviewed methods to assess progression of OA of the hip and knee [2]. At the request of the U. S. Food and Drug Administration, an independent committee has developed a set of guiding principles for the development of new drugs for OA [3]. Subsequently, the European Group for the Respect of Ethics and Excellence in Science (GREES), through a subcommittee, has made recommendations regarding the methods to be used for registration of drugs for OA [4]. Most recently, the Outcome Measures in Arthritis Clinical Trials (OMERACT) group has recommended a core set of measures to be used in OA clinical trials [5].

The Osteoarthritis Research Society also established a Task Force to address the issue of clinical trial guidelines for OA. Through a series of meetings, a draft manuscript was developed. The intent of the Task Force was to bring together the ideas on the conduct of clinical trials generated by the relevant active working groups, and to add sufficient detail to be of help to any party involved in the design of clinical trials. The Task Force was composed of academic and clinical physicians, researchers in the pharmaceutical industry and members of GREES. Representatives of regulatory agencies were invited to attend all meetings.

On May 26 and 27, 1996, a Workshop attended by representatives of the basic and clinical sciences, the pharmaceutical industry, GREES, and regulatory agencies was held in Washington, D. C. to discuss the working document of the Task Force. The present document resulted from the Workshop and reflects a consensus of the participants (See Appendix I).

It can be expected that the metrology and methodology of clinical trials of drugs for OA will change in the future, as they have in the past [6, 7]. The following recommendations for the design of clinical trials in patients with OA are made with the understanding that they will require modification as new information becomes available. Investigators, regulatory and sponsoring agencies should be aware of the likelihood of such changes. Investigators and sponsors will need to incorporate new methodologies into their protocol design, and regulatory agencies will require flexibility to adapt to the newer technologies and methodologies. Indeed, as part of the advancement of science, it is expected that OA protocols will contain both validated measures and investigational outcome measures still requiring validation. The following are recommendations, or guidelines, not rigid rules for the conduct of clinical trials in OA. Many of the recommendations are supported by published clinical research. However, some recommendations have yet to be validated and are based on the best judgment of the Task Force and the participants of the Workshop.

B. Objectives for treatment of OA

Medications for OA may affect symptoms and/or modify structure (joint pathology). Demonstration of these benefits will depend upon the trial design and outcome parameters selected. Trial design will depend on the mechanism of action of the drug and the expected response.

For trials related to symptoms, some measure of joint pain will usually be the primary outcome variable. Factors that are considered in trial design include, but are not limited to, the pharmacodynamics of the drug, time to clinical response, duration of benefit after discontinuation of treatment, route of administration, frequency and severity of adverse events, effects on pain, effects on inflammation and effects on other symptoms and signs of the disease. In contrast to a prior consensus publication [1], the majority of the members of the Task Force and participants in the Workshop felt that there is no advantage in creating a separate class for those agents that produce a rapid symptom response from those with a slower onset of benefit. Medications used to treat symptoms have generally included analgesics and nonsteroidal anti-inflammatory drugs (NSAIDs). Examples of agents that may prove to be of benefit with a particularly prolonged onset to pain relief include intra-articular (IA) hyaluronic acid, oral glucosamine, chondroitin sulfate and diacerrhein. For the purpose of this report, the term symptom modifying drugs for OA will be used for both rapid and slow onset agents.

A drug may have effects on joint structure/function independent of its effects on symptoms.

Studies of drugs that are expected to modify the pathologic process of OA should measure outcome parameters that reflect an alteration of joint structure. Such drugs may (1) prevent the development of OA, and/or (2) prevent, retard, reverse, or stabilize the progression of established OA by altering the underlying pathologic process(es). A drug that affects the pathology of OA may have no direct effect on joint symptoms. Symptomatic improvement may occur only after a prolonged period of administration. Demonstration of symptomatic improvement is not required if no claim is made for this outcome. Indeed, the GREES group have clearly separated those drugs that may alter the structure without an affect on symptoms from those that modify structure and do effect symptoms [4]. Whether related to symptoms, function or some other variable, the primary outcome measure should be clinically relevant.

Drugs with a potential for structure modification have been labeled as 'chondroprotective', disease modifying drugs for OA (DMOADs), 'anatomy modifying agents', 'modifiers of morphology', etc. There is no uniformity of opinion concerning the term that best reflects the action of these agents. For the purposes of this report, and to provide consistency in the literature [4], the term structure modifying drugs will be used. To date, no agent has been proved to have structure modifying properties in humans. It should be pointed out that a symptom modifying drug may prove to have structure modifying properties (favorable or deleterious), just as a structure modifying drug may have symptom modifying properties.

C. Levels of clinical trials for OA

Preclinical studies are helpful in assessing potential modes of action and the dose range for benefit/toxicity, and may shorten the duration of clinical testing of a potential structure modifying drug. Although they are not essential, studies that demonstrate efficacy in animal models of OA will strengthen the rationale for clinical trials of structure modifying drugs in humans.

Medications undergoing clinical investigation are allocated to different levels of development as described below [8].

C.1. PHASE 1 TRIALS

Phase 1 trials are directed principally at demonstrating pharmacokinetics and safety. They may also contain a dose-finding component. Escalating dose trials are desirable for initial evaluation of drug safety. Mechanism based pharmacological evaluations, including those at the site of action (i.e., in joint tissues), are common. Initially, the presence of comorbid conditions should be minimized: later studies may target special populations, such as individuals taking concomitant medication. Phase 1 trials may be performed in normal volunteers or in a patient population appropriate for the target indication. Double-blind, placebo-controlled, single and multiple dose Phase 1 trials are desirable for the initial evaluation of drug safety. Evaluation of efficacy is not the primary purpose of Phase 1 trials. A Phase 1 trial cannot adequately address the benefits of structure modifying drugs.

C.2. PHASE 2 TRIALS

The goals of Phase 2 trials are to define an ideal effective dose range and regimen (Phase 2 trials must take into account both drug activity and toxicity) and to provide sufficient patient exposure to demonstrate safety in order to justify progression to Phase 3 trials (See Below). The duration of the study and number of patients studied should be based on the mechanism of action of the drug, duration of action of the drug, outcome variable being assessed, variability of the outcome parameters, and the intended patient population. Dose ranging in these and subsequent studies should identify the minimal effective dose and dose-response profile, and may define the maximum tolerated dose of the drug in patients with OA.

C.2.1. Symptom modifying drugs

Phase 2 studies of symptom modifying drugs for OA should be placebo controlled, randomized and double-blind. Efficacy can often be demonstrated within days. Longer studies (weeks) are needed to demonstrate slow onset or persistent benefit. Even longer studies are required for safety. In studies of long-duration, rescue analgesia may be necessary. A short-acting analgesic is suggested with a suitable washout employed prior to efficacy assessment.

C.2.2. Structure modifying drugs

As an alternative to demonstrating effects on joint structure, dose-ranging studies in Phase 2 trials of a structure modifying drug may utilize other measures of mechanism-based drug activity. Because these are measures of physiology and not efficacy endpoints, multiple dose regimens may be

needed in late Phase 2 (2b) or Phase 3 trials. The duration of Phase 2 studies for a structure modifying drug will also depend on its mode of action.

C.3. PHASE 3 TRIALS

Phase 3 trials are intended to convincingly demonstrate efficacy and safety of the optimal regimen and dose(s) of the test agent. Replication of pivotal studies (studies of primary importance for registration of drugs) for demonstration of efficacy is recommended. There should be only one target joint in a single trial. These studies are designed to clearly define the dose/regimen of the test drug to be recommended for clinical use, further define toxicity, and compare the test drug with a reference drug and/or placebo. Sample size and study duration should be calculated to assure that subjects will be followed for a sufficient time period to detect a clinically relevant, as well as a statistically significant, difference between treatment and control groups with respect to efficacy—outcome parameters (see Statistical Methods). Sufficient data must be supplied to the appropriate regulatory agency(ies) to satisfy safety concerns. The number of patients and length of time to assess safety should follow the recommendation for chronic diseases of the *Guidelines for Industry* [9].

C.3.1. Symptom modifying drugs

Phase 3 trials of drugs with a rapid onset of effect can be as short as 4 weeks. At times, shorter trials are appropriate. Longer trials may be needed to evaluate efficacy for drugs with a slower onset of action. In studies of long-duration, rescue analgesia may be necessary. A short-acting analgesic is suggested with a suitable washout prior to assessment of efficacy. A Phase 3 double blind study may be followed by a long-term double-blind study or open-label extension to evaluate safety.

C.3.2. Structure modifying drugs

There are no proven structure modifying drugs. Hence, the extent of testing needed to demonstrate this effect is not established. The duration of the trial should be predetermined, and it is recommended that it be at least 1 year. The duration will depend on the mode of action of the drug, the anticipated response rate, the primary outcome variable and the length of time needed to show a difference in comparison with a control (i.e., placebo) group. Structural changes are required as primary endpoints. The size of the study population be should be ideally calculated on the basis of preliminary data from Phase 2 trials in the particular population to be studied (see Statistical Section).

C.4. PHASE 4 TRIALS

Phase 4 studies are performed after the agent has been approved for clinical use by the regulatory agency. These studies may be used to support clinical observations leading to expanded indications. They also permit exploration of uncommon adverse events that can be discovered only in studies with a large sample size. It also provides supportive evidence of long-term benefit. Some Phase 4 trials may be open label. To date, Phase 4 trials have been published only for symptom modifying drugs.

C.5. REGULATORY ISSUES

When evaluating OA medications, it is advisable (when applicable) for the sponsor to schedule a pre-investigational new-drug meeting with the appropriate regulatory agency to define the preclinical and clinical requirements prior to initiation of Phase 1 trials. The sponsor should maintain communication with the regulatory agency as the drug progresses through Phase 2 and Phase 3 studies.

D. Entering patients in OA trials

This section addresses several aspects of the study design, including the protocol, admission criteria, selection of the study population and the definition of what is to be studied. Baseline assessment should provide information on joint localization (site), etiology (primary, secondary), severity of symptoms, structural abnormality in the joint, concomitant therapy and comorbidity [4].

D.1. OVERVIEW OF THE PROTOCOL

The study protocol should be divided into sections that encompass background information, rationale for the study, the question(s) being asked, size and site(s) of the study, method of patient selection (including inclusion and exclusion criteria), the method of procedure, clearly defined primary and secondary outcome variables, specific measures to be performed at each visit, drug dispensing format, method of reporting adverse events, statistical analysis and regulatory issues (including drug accountability, institutional requirements, etc.).

It is desirable to include a table (or flow sheet) that outlines the method of procedure, information from selected references (e.g., disease classification, radiographic criteria), the informed consent statement, protocol worksheets, drug accountability forms, the data collection forms, etc.

The protocol should carefully define the investigators, their study sites, the method of randomization, patient monitoring procedures, technical aspects of imaging techniques, laboratory tests, methods of documenting adverse events, methods of blinding and method of documenting medication intake for each patient (active drug, placebo, rescue analgesia), and the method of maintaining the medication log for each participating center.

D.2. DEMOGRAPHICS

Demographics recorded in the protocol should include identifying information, such as the patient's name, address and telephone number, which should be kept confidential. The patient's name should be coded by letters/numbers for data processing and future reference.

As a minimum, sociodemographic and clinical data collected at the time of enrollment into the study should include age (date of birth), sex, race, height, weight, marital status and years of formal education.

D.3. DIAGNOSIS

Criteria for diagnosis of OA should be clearly stated. Patients should fulfill validated criteria for the classification of OA, such as those published by the American College of Rheumatology (ACR) [10–12]. The disease should be classified as primary or secondary. Study populations should be as homogenous as possible with regard to the presence of idiopathic (primary) or secondary OA [10]. If patients with secondary OA are studied, the underlying condition should be specified and should be the same in all patients (e.g., post-traumatic arthritis, mechanical derangement of the knee). It is suggested that in studies of patients with idiopathic OA, exclusions for secondary OA of the study joint include septic arthritis, inflammatory joint disease, gout, Paget's disease of bone, recurrent pseudogout, articular fracture, major dysplasias or congenital abnormality, ochronosis, acromegaly, hemachromatosis, Wilson's disease and primary osteochondromatosis [4].

D.4. RADIOGRAPHS

The radiographic severity of OA in each patient should be quantified and documented using either aggregate radiographic criteria (e.g., Kellgren and Lawrence scale [13, 14]) or grading of specific radiographic features [15–17]. This estimate of anatomic alteration on images should be acquired within 3 months after entry. The range of grades used for entry criteria, as well as variations in grade among treatment and placebo (or control) groups should be comparable and similar. These radiographic entry criteria should also be appropriate for the specific study design. For example, a cohort that included advanced severity might be appropriate in studies of a symptom modifying drug while a cohort limited to minimal severity would be more appropriate for studies of a structure modifying drug intended to retard progression.

D.5. STUDY POPULATION

The source of the patient population (e.g., clinic-based, community-based, hospital-based) should be defined in the protocol. Considerable controversy exists regarding the use of broad vs narrow patient eligibility criterion. Broad patient eligibility allows for generalizable application of positive results; however, because of the larger amount of variation, broad patient eligibility increases the sample size of the study population required to demonstrate clinical and statistically significant differences, and may mask the presence of subsets receiving benefit (unless extensive stratification is performed). At the Workshop, the consensus was that patient eligibility should define specific populations and that, where appropriate, stratification of subgroups should be employed within studies for secondary endpoints of interest.

Examples of high-risk groups that might be considered for inclusion in studies of structure modifying drugs include obese women with unilateral radiographic knee OA [18], and men or women who have undergone meniscectomy [19, 20]. Examples of variables to be considered for stratification of the source population might include prior surgical intervention of the index joint, and high- vs low-risk groups. Examples of subjects who might be considered for exclusion might encompass either low- or high-risk populations, such as young age (<45 years old) and those with protrusio acetabuli, concentric femoral head migration, extensive surgery of the reference joint, excessive varus/valgus deformity, concomitant rheumatic illness (e.g. fibromyalgia), and those

involved with litigation/compensation related to the reference joint.

D.5.1. Symptom modifying drugs

For studies of symptomatic response, the level of symptoms at baseline should be of sufficient severity to permit detection of change, i.e., not too mild. After washout (see Section E.5.), inclusion criteria for symptomatic response should include the following:

- Pain of at least mild intensity: e.g., 100 mm visual analog scale (VAS) recording of $\geq$25 mm; or five point categorical (Likert) scale grade $\geq$1 (where 0 is no pain and 4 is extreme pain);
- Definite radiographic changes of OA, using an established scale and atlas, e. g., Kellgren and Lawrence radiographic grade $\geq$2 for tibiofemoral OA (i.e., presence of a definite osteophyte); modified Croft scale $\geq$2 for hip OA [13, 14, 21, 22].

D.5.2. Structure modifying drugs

For studies of structure modifying drugs, as discussed above, special subpopulations of subjects who are at high risk for development of OA or rapidly progressive OA may be advantageous (as above). In addition, the following should be considered:

- Kellgren and Lawrence radiographic entry criteria: prevention studies: grades 0 or 1 (i.e., absence of a definite osteophyte); disease retardation/reversal studies: grades 2 or 3 (i.e., sufficient remaining interbone distance to permit detection of worsening/progression);
- Current or previous pain in the index joint is not essential. However, changes in pain may be examined as a secondary outcome measure

Preliminary data suggest that some molecular markers in serum may predict radiographic progression of established OA [23, 24]. Analysis of molecular markers may select subpopulations who are most likely to show progression in OA.

D.6. INCLUSIONS/EXCLUSIONS

Inclusion criteria should be clearly defined and should specify the population to be studied by age, sex, diagnostic criteria, joint with OA, degree of symptoms, and radiographic grade.

Exclusion criteria should similarly be clearly defined with regard to degree of symptoms, radiographic grade, concomitant disease, prior peptic ulcer disease (if a drug is perceived to have gastrointestinal effects), concomitant medications, pregnancy/contraception, IA depocorticosteroid or hyaluronic acid injection, tidal lavage, secondary OA (listed above).

Opinion varies concerning the proximity to the beginning of a study for administering IA medication into the reference joint. All agreed that there should be a sufficient interval between the time of the injection and the beginning of the study to eliminate the confounding effects of the injection on joint pain. The consensus of the participants at the Workshop was that a minimum of 3 months should elapse between the time of the IA injection and the trial (e.g., IA corticosteroids). This interval may be longer for specific types of IA therapy (e.g., IA hyaluronan), but sufficient evidence is not available to provide more definitive guidance at this time. The investigator should consider stratification of patients receiving prior IA therapy administered within a year of the study.

Additional exclusions are significant injury to the affected joint within 6 months of trial start; arthroscopy of the affected joint within 1 year; disease of spine or other lower extremity joints of sufficient degree to affect assessment of the target joint, use of assistive devices other than a cane (walking stick) or knee brace, concomitant rheumatic disease (e.g., fibromyalgia), or poor general health interfering with compliance or assessment.

As with any investigational drug, women of childbearing potential should be screened for pregnancy, and if pregnant, should be excluded from the trial.

D.7. OA HISTORY

The OA history is used to characterize the study population and should include the location and number of symptomatic OA joints; presence of hand OA (e.g., Heberden's nodes in patients with hip or knee OA); duration of symptoms; duration of the diagnosis of OA; history of prior medications for OA; surgical procedures performed on the study joint (including arthroscopy), with the date of the most recent procedure; use of assistive devices, such as canes, crutches, knee braces (in studies of lower extremity OA); history of prior IA (e.g. depocorticosteroid or hyaluronan) injection, with date of most recent injection (see above).

D.8. HISTORY (OTHER)

Other baseline history that may be of value includes smoking history, hormonal status in postmenopausal women, concomitant chronic disease, and concomitant medications, e.g., estrogens, anti-inflammatory drugs.

D.9. STUDY JOINT

Protocols should be limited to the evaluation of a single joint site (e.g., knee, hip) or in the case of hand OA, either both hands or the symptomatic hand (preferably the dominant hand).

D.9.1. Symptom modifying drugs

Although data may be collected for both right and left joints (e.g. knee, hip), for symptom studies only one should be the primary joint evaluated (except for hands as above). This is most often the signal (more symptomatic) side. Changes in the contralateral joint should be considered as a secondary outcome variable.

D.9.2. Structure modifying drugs

For studies of a structure modifying drug, the more involved side of a single joint site (e.g., hip, knee) should be studied as the primary outcome variable. In these cases, changes in the contralateral joint can serve as a secondary outcome variable. However, changes in the contralateral joint, which may not yet be symptomatic or have definite OA, may be selected as the primary outcome variable (e.g. Chingford data) [18].

For studies of both symptom and structure modifying drugs, additional joint sites may be evaluated as secondary outcome variables.

D.10. PHYSICAL EXAMINATION OF THE INDEX JOINT

Baseline information about the index joint helps characterize the study population and provides reference data for assessing how variables of interest have changed during the course of treatment. Evidence of inflammation (e.g., joint effusion), joint deformity, and joint contractures should be noted. For large joints, loss of range of motion and presence of severe valgus/varus deformity may be useful as exclusion criteria. Although it is important to record the presence of clinical signs of inflammation, including synovial effusion, these should not be used as a primary outcome measure in trials of structure modifying drugs.

D.11. FUNCTION

Measuring the degree of functional impairment can identify the severity of disease in the study population. Functional impairment should be defined using a segregated, validated multidimensional index (SMI) such as the Western Ontario and McMasters Universities (WOMAC) [25] OA index for hip and knee OA, or an aggregated multidimensional index (AMI) such as the Algofunctional Index (AFI) for hip or knee [26]. At this time, although the AFI has been validated, separate pain, stiffness and physical function subsections have not been validated for independent application.

D.12. GENERAL PHYSICAL EXAMINATION

A general physical examination should be performed at the onset of the study and again at the end of the study.

D.13. INFORMED CONSENT

Guidelines for information to be contained in the Informed Consent statement should be in accordance with the Declaration of Helsinki [27]. Patient participation requires understanding, and completion of an informed consent document that has been approved by the appropriate institutional review board.

E. Conduct of the study

This section deals with the procedures used during the study, exclusive of individual outcome variables.

E.1. STUDY DESIGN

Studies should generally be single joint, controlled, randomized, double-blind, and parallel in design. Occasionally, crossover studies or other designs may be appropriate.

The study should include a screening and baseline visit. The two visits allow the collection of more reliable baseline data, assure that the patients fulfill entry criteria and may be used to help reduce noncompliance ('faintness-of-heart test' [28]), collect biological specimens, etc. For treatment group assignment, patients should be randomized in the order in which they are enrolled into the study, to receive treatment according to a randomization schedule specifically designed to meet study objectives.

At each visit, vital signs (blood pressure, pulse,

and weight) should be recorded and a report of adverse experiences (see below) obtained.

In order to minimize unwanted sources of variation in patient assessment, to the extent possible, the same examiner should examine the same patient at each visit, at the same time of day (and preferably also on the same day of the week) throughout the duration of the trial.

E.2. PRIMARY STUDY OUTCOME

Efficacy studies of OA drugs should preferably identify a single clearly defined primary outcome variable. The choice of this variable will depend upon the nature of the desired drug effect and the objective of the study.

An alternative approach might involve the use of several primary outcome variables. With this latter approach, adjustments to the significance level are required for multiple analyses performed. (See Outcome Measures below.)

E.3. SECONDARY STUDY OUTCOMES

The inclusion of one or more secondary outcome variables will strengthen the study design. Collection of information for the secondary outcome variables should not interfere with collection of data for the primary outcome variable.

E.4. EXAMINER

The method used for training and masking of the examiner and masking of the patient must be specified. Both a blinded investigator (to assess the patient for efficacy and adverse events) and an unblinded investigator may be needed to administer the test medication and monitor toxicity in some studies.

E.5. WASHOUT REQUIREMENTS

E.5.1. Symptom modifying drugs

All symptom-oriented studies require discontinuation of prior analgesic and anti-inflammatory medications, including topical agents, prior to initiating treatment with the test drug in order to permit an evaluation of unmodified pain severity. The time of withdrawal should be the time required for the clinical effect to disappear (e.g., 5 half-lives of the drug). During the washout period, subjects may use acetaminophen (or paracetamol) as rescue analgesia (up to 4 g/day in the U.S. and up to 3 g/day in Europe). This must be discontinued in sufficient time for the clinical effects of the rescue drug to disappear.

Worsening of symptoms during the washout period—although not necessarily a requisite for subject inclusion into the trial—should be documented.

E.5.2. Structure modifying drugs

A washout period is not required in trials of a structure modifying drug. If however, the effect of the drug on symptoms is to be tested, then the use of a washout period should be considered.

E.6. ADMINISTRATION OF STUDY MEDICATION

Control agents may include placebo or active (e.g., analgesic or NSAID) agents. Use of placebo may be influenced by ethical and regulatory agency considerations. Active control agents offer the advantage of demonstrating improved efficacy over existing therapies, but may require large numbers of subjects.

E.6.1. Topical

Topical test medications should be dispensed in containers which are identical in appearance to those containing the comparison agent (drug or placebo). The comparison agent should mimic the test medication in appearance, odor and local effects on the skin. Clear instructions regarding use must be provided to the patient both orally and in written form and must be contained in the Informed Consent. Compliance should be monitored by weighing the returned tubes or measuring the returned liquid. Placebo responses are particularly frequent with this technique of drug delivery, so placebo controlled trials are particularly important, as are carefully defined, homogeneous study populations.

E.6.2. Oral

Oral test medications should be formulated to provide an appearance identical to that of the comparison drug (placebo or other). If this is not feasible, a 'double dummy' technique (two non-identical active agents, each with an identical matching placebo) should be used.

Preferably, medication should be dispensed in blister packs with the label clearly stating the day and time of administration. Compliance should be monitored by counting returned unused medications or by use of medication vials with computerized caps.

Concomitant medication (e.g., rescue analgesia and NSAIDs in studies of structure modifying drugs) may be dispensed in bottles. The pills should be counted at each visit. Analgesic drugs with a short half-life should not be taken from the evening prior to the day of the evaluation if pain is to be evaluated.

E.6.3. Parenteral medication

Parenteral medication should be formulated to provide an appearance identical to that of the comparison drug. If this is not possible, the parenteral medication should be dispensed by a person other than the blinded investigator (e.g., by an unblinded investigator) and the injectable agent should be concealed from both the patient and the blinded evaluator.

E.6.4. IA Medication

IA study medication should be formulated to provide an appearance identical to that of the comparison drug. If this is not possible, the medication should be injected by a physician other than the blinded investigator (e.g., unblinded investigator). The volume of control (carrier) injected should equal the volume of the test agent. The joint should be aspirated to remove any existing effusion as completely as possible prior to instillation of the drug, and the volume of fluid removed should be recorded. The injectable should be concealed from both the patient and the blinded evaluator. Placebo responses are particularly frequent with this technique of drug delivery, so placebo controlled trials are particularly important, as is the use of carefully defined homogeneous study populations.

E.7. COMPLIANCE AND SUBJECT RETENTION

It is essential for studies of structure modifying drugs, that strategies be employed to maximize and document patient compliance. For example, contact might be maintained with patients at 4–8 week intervals by telephone. The method of communication and time spent with patients should be standardized as much as possible without jeopardizing the relationship with the patient.

E.8. SOCIOECONOMIC MEASURES

Sponsors should consider performing pharmacoeconomic analyses in all OA clinical trials [29, 30].

E.9. USE OF CONCOMITANT MEDICATIONS

E.9.1. Symptom modifying drugs

It is impractical to expect patients to participate in a long-term trial without some potential for use of rescue medications for pain. For long-term trials, use of concomitant medication should be permitted on a limited basis. An example may be the use of acetaminophen (or paracetamol) for escape analgesia (up to 4 gm/day in the U.S. and up to 3 gm/day in Europe). Any escape medication must be discontinued in sufficient time for the clinical effects of the agent to disappear prior to the assessment. Protocol design should include a record of the consumption of analgesics, NSAIDs, and IA injections. However, the use of such information as an outcome in clinical trials has not been validated.

IA depocorticosteroids should not be permitted in studies of symptom modifying drugs, except as part of the protocol design.

E.9.2. Structure modifying drugs

Concomitant therapy may interfere with the evaluation of outcome measures and should ideally be excluded. However, in long-term studies, it is neither ethical nor practical to exclude all concomitant treatments. In all trials, concomitant therapies (drugs or other interventions) that are likely to affect joint structure should be excluded, and rescue therapy should be standardized, carefully recorded and monitored. As noted above, participants may use acetaminophen (or paracetamol) for escape analgesia (up to 4 g/day in the U.S. and up to 3 g/day in Europe). Analgesics and NSAIDs must be discontinued prior to the assessment in sufficient time for the clinical effects of the rescue medication to disappear.

The consumption of analgesics, NSAIDs, and IA injections should be documented at each visit. However, methods need to be developed to effectively control for these confounding variables in the analysis and the use of this information has not been validated as an outcome variable.

E.10. CONCOMITANT NON-MEDICINAL THERAPY

Concomitant treatment with physical and/or occupational therapy should be either standardized or adjusted for in the analysis to ensure that the effects of exercise programs on disease progression do not bias the outcome of the study. Information on weight change (reduction or gain), changes in use of ambulatory support (cane,

crutches, walker), and introduction of, or changes in, physical or occupational therapy during the study should be incorporated into the study design.

E.11. LABORATORY TESTS

For most multicenter studies, routine laboratory tests (complete blood count, urinalysis, serum chemistry determinations) should be performed in a central laboratory.

Routine synovial fluid analyses should be performed at each site, and should include an examination for cells and crystals.

For studies routinely performing arthrocentesis with injection of an IA agent, culture of the synovial fluid should be performed as clinically indicated.

E.12. ADVERSE EVENTS

Adverse events should be ascertained in an open-ended manner, rather than by checklist. They should be recorded at each visit and between visits, as appropriate. The date of onset, severity, a judgment with respect to the relationship between the adverse event and the test agent, treatment and the duration, and resolution of the adverse event should all be recorded.

Serious adverse events should be reported to regulatory authorities immediately.

E.13. PROTOCOL VIOLATION

Reasons for termination of a subject from the study due to protocol violation must be specified in the protocol. Intake of rescue medications (other than those specifically prescribed), use of oral or topical agents, or devices targeted toward pain relief during the course of the study should be prohibited. Information on the use of such agents should be obtained at each visit and recorded, and the patient should be warned about such co-interventions. Patients in repeated violation of the protocol may need to be dropped from the study.

Screening for protocol violations by performing blood or urine analyses for salicylates or related agents is not considered useful.

E.14. CASE REPORT FORMS AND SUPPLIES

Investigators must maintain adequate records showing the receipt, dispensing, return, or other disposition of the investigational drug, including the date, quantity, batch or code number, and identification of subjects who received the study drug. Investigators must maintain completed case report forms and informative source documents. Case report forms must be kept in locked cabinets to maintain security. There are no special requirements for OA trials.

F. Outcome measures of OA

Instruments used to measure outcome in clinical trials of OA should be valid, reliable and responsive to change, when such measures exist. Clinical trials in OA should use published instruments that have been used in other studies, thus permitting comparison of results across trials of different therapeutic interventions. Clinical trials in OA should include a core set of validated measures [5] (Appendix II):

- Pain
- Physical function
- Patient global assessment
- Imaging (for studies ≥ 1 year in duration)

Additional measures that are recommended include the following:

- Quality of life/utility (strongly recommended)
- Physician global assessment

Optional measures for trials in OA include the following:

- Signs of inflammation
- Biologic markers
- Stiffness
- Performance based measures of function
- Presence of 'flares'
- Time to surgery
- Analgesic consumption

The items listed below pertain mostly to phase 3 trials. These measures should be recorded at baseline and serially at appropriate intervals.

F.1. SYMPTOM MODIFYING DRUGS

For studies of drugs designed to affect symptoms, the primary outcome variable should usually be joint pain reported by the patient. Measurement should be serially recorded at appropriate intervals, at least monthly. However, this is dependent upon the target joint and study design.

F.1.1. Pain

The degree of joint pain in the index joint(s) should be graded. Pain should be recorded on a five-point Likert scale (e.g., none, mild, moderate, severe, very severe) or on a 100 mm VAS. Single questions about pain can be used but the activity

causing pain should be specified: e.g., weight bearing, resting, nocturnal, post exercise, stair climbing. Alternatively, a validated pain instrument can be used (e.g., WOMAC pain subscale [24]). Other pain indices include the Health Assessment Questionaire (HAQ) [31] and Arthritis Impact Measurement Scale (AIMS) [32].

F.1.2. Function

The AFI [25] and the function subscale of the WOMAC [24] have been validated and are recommended for studies of OA of the hip and knee. Other indices which have been used include the HAQ disability index [28], and AIMS [33]. Disability indices specifically designed to measure hand function are under development [34, 35].

F.1.3. Global status

F.1.3.a. Patient assessment of global status. The patient's assessment of his/her global status should be measured using a Likert or VAS scale. The optimal method by which this should be measured is not well established. However, a standard question should be asked, e.g., 'Considering all the ways your OA (joint site) affects you, how are you doing (time frame)?'

F.1.3.b. Physician assessment of global status. A measure of the physician assessment of global status may be required by some regulatory agencies. There is no generally accepted method for measurement of this variable. A question such as 'Considering all information, how is the patient's OA [joint site] today?' should be used with a VAS or Likert scale.

F.1.4. Quality of life scales

Measurement of health-related quality of life and utility based measures at appropriate intervals is strongly recommended; although, these are not a part of the core set of measures. Examples of health related quality of life instruments include the Medical Outcomes Study, 36 question short form (SF-36) [36], Sickness Impact Profile (SIP) [37], Nottingham Health Profile (NHP) [38], and EuroQol [39]. Examples of utility instruments include the Time Trade Off, the Standard Gamble and Techniques and Feeling Thermometer and the Health Utilities Index (HUI) [40, 41].

F.1.5. Joint examination

Measures of range of motion, intermalleolear distance, knee interbone distance, heel to buttock measurements, knee circumference, etc. have been validated to variable degrees [42]. The usefulness of these measures in clinical trials remains unclear and their inclusion is optional.

F.1.6. Performance-based measures

Performance-based measures which include such items as grip strength, time to walk a specified distance (e.g., 6 or 15 m, 50 ft), distance walked in a specified time (e.g., 6 min)., have been studied to a variable degree. Some composite measures exist [43]. The usefulness of these measures in clinical trials remains unclear and their inclusion is optional.

F.1.7. Inflammation

Clinimetric properties of methods designed to measure inflammation have not been well elucidated. The usefulness of these measures in clinical trials remains uncertain.

F.1.8. Response criteria

There is no definition of a minimum clinically important response for the above measures. Available information does not allow setting of predetermined limits for improvement. This is particularly true for the composite indices. At this time, the Task Force recommends that each protocol predefine a significant response, based upon statistically significant improvement in a carefully defined primary efficacy variable (see the statistical section below). At this time, the Task Force does not recommend using an individual response criterion such as has been recommended in rheumatoid arthritis [44].

F.2. STRUCTURE MODIFYING DRUGS

For studies of potential structure modifying drugs, the primary outcome variable should be a measure of joint morphology; e.g., imaging (see below) or direct visualization, i.e., arthroscopy. As stated above, time to joint replacement surgery is not recommended as a primary outcome variable due to its dependence on factors unrelated to disease progression. Clinical follow-up of patients participating in trials of structure modifying drugs should be at intervals of 3 months or less.

F.2.1. Radiography

The primary radiographic evaluation should be of a single joint (knee, hip, hand). Outcome should assess the effect of the drug on joint structure. Although assessment should include both cartilage and bone, the primary radiographic outcome variable for studies of progression of the hip and knee should be minimum joint space width (JSW), since this measure is more sensitive than global scoring [45–49]. Osteophytes and other bone changes should be assessed as secondary outcome variables either by measurement or by grading, using published atlases [13–17]. In contrast, for studies of prevention, the primary radiographic outcome variables should include osteophytes, since this feature is most strongly associated with knee pain, is a basic component of the ACR classification criteria, and is the hallmark of the Kellgren–Lawrence scale of the knee. Outcome variables for hand OA should be based on published atlases agreed upon by the study group in advance.

Obtaining reproducible X-rays on successive visits is a prerequisite for reliable assessment of progression of OA. The sources of variability in joint space width measurement are numerous (patient positioning, radiographic procedure, measurement process, etc.), protocols have been proposed for hip and knee joints [50–52]. It is essential to standardize radiographic technique based on published, validated data (Appendix III). The method should define the radio-anatomic position of the joint, beam alignment, and should define the anatomic landmarks for measurements. Positioning of the patient should also be based on validated published methods, but in all cases, weight bearing (standing) anteroposterior views should be used in studies involving the hip or knee. Repositioning of the joint can be facilitated by use of foot maps drawn at the time of the initial examination. Correction for radiographic magnification has been shown to improve accuracy and precision of measurements [48, 53]. Techniques that improve the precision of measurements might lead to studies requiring smaller sample sizes.

Quality assurance should include training sessions for technologists at the onset of the study as well as for any technologists recruited during the study. Radiographic quality, including patient positioning, exposure, labeling, etc., should be monitored throughout the study. Even minor changes in technique may significantly alter the precision of measures of joint anatomy and hence conclusions about treatment response. It is, therefore, critical that the technique be identical at all centers involved in a multi-institutional study and remain consistent throughout the study.

The number of readers, method of blinding and the method of manual measurement should be agreed upon in advance by the study group. Quality control of the readings should include an initial training session and periodic assessments of performance. Validated methods for computerized reading of digitized radiographs can decrease observer-based error. Enhanced anatomical detail provided by microfocal magnification radiography can further improve precision and accuracy of measurements [48, 54].

F.2.2. Magnetic resonance imaging (MRI)

MRI is uniquely capable of visualizing all components of the joint simultaneously, and therefore offers an opportunity to asses the joint as an organ. MRI is capable of quantifying a number of morphological and compositional parameters of articular tissues relevant to OA. Recently developed techniques for noninvasively quantifying cartilage volume, thickness and water content, particularly in early disease, show promise as potential outcome measures for future therapeutic studies (Appendix IV). While some cross-sectional measures have been validated, their performance in longitudinal studies has yet to be determined.

F.2.3. Other imaging modalities

Computed tomography, ultrasonography and scintigraphy have not been adequately validated and cannot be recommended for use in long-term studies.

F.2.4. Arthroscopy

Arthroscopy can directly visualize cartilage and other IA structures, including fibrocartilagenous menisci, synovium, ligaments and chondrophytes. Attempts to quantify this information have followed two strategies. The first transforms information from each cartilage lesion into a numeric score, weighted mainly by depth and size of the lesion. When several lesions are found, as occurs frequently in OA, a composite score is derived from the scores of individual lesions. The second approach calls for the arthroscopist to globally assess cartilage degeneration in a compartment-by-compartment fashion, recording each impression on a VAS. Both strategies are being employed in the two systems currently under evaluation, with intra- and inter-observer reliability determined for both [55, 56], and

sensitivity to change (utilizing videotaped records from two points in time) has been shown for the French system [51].

Other systems yet to be devised may prove superior for assessment of particular aspects of OA, examining biomechanical characteristics of cartilage (which might be shown better by a probe) or features of the accompanying synovitis. The precision and sensitivity to change of any system employed in an OA outcomes trial should be determined by a study group before the system is implemented. Management of arthroscopic data by videotaping each procedure provides an immutable record that can be reviewed by a blinded evaluator. However, video records do not convey certain impressions obtained in real time, such as three-dimensional perception and tactile feedback from probing the cartilage. Regardless of the recording technique, a systematic uniform method of collecting arthroscopic data is essential, and should be specifically delineated in any protocol. Discussion of the technical aspects of the arthroscopic procedure is beyond the purview of this report. However, the size and type of instrument used and conditions under which the procedure is performed should be uniform for all investigators in any particular study.

F.2.5. Molecular markers

Molecular markers have not been validated as outcome measures in clinical trials of OA (Appendix V). However, molecular markers have the potential of offering a unique way of assessing drug effects on specific disease mechanisms, and modes of action of drugs in phase 1 clinical trials [57–59]. The field is developing rapidly. For these reasons, trials should include collections of body fluid samples. Standardization of methods for collection and storage is important.

G. Statistical methods

There are specific statistical tasks in the design, implementation and analysis components of a clinical trial. General textbooks cover a broad range of topics regarding statistics in clinical trial research [60, 61].

G.1. DESIGN

The predominant activity of the statistician is working with the researcher in developing the protocol. The protocol must clearly list the primary and secondary study objectives. Where appropriate, these objectives should be rephrased as null versus alternative hypothesis to be tested.

All protocols should specify the outcome measure(s) to be used for evaluating the study treatments and should contain sample size calculations for all primary outcomes, indicating the required number of patients to achieve prestated power and significance levels, or a calculation of the power provided with a prestated sample size. Sample size calculations are based on the choice of experimental design (e.g. parallel groups, factorial design, more than one treatment group vs control) and require that explicit assumptions be made regarding the variance(s) in outcomes among study subjects and the desired magnitude of change(s) in the outcome variable(s) during the study period; these assumptions should be stated explicitly in the protocol. Phase 3 studies should require a 5% or lesser level of significance and 80% or greater power to detect a protocol defined minimal clinically meaningful difference in the expected outcome between the treatment and control groups. These assumptions should, when possible, be based on available clinical/epidemiological data.

Randomization is a method for assigning patients to a test or control treatment that is free of selection bias. The method for randomization should be specified in the protocol. Two general designs exist for randomization of patients to treatments: fixed randomization and adaptive randomization. Fixed randomization schemes may be completely random or may be constrained so as to ensure balance in the number allocated to various treatment groups (randomization in blocks of fixed size, stratified random sampling). Randomization in blocks should be considered if patient enrollment is likely to continue over an extended period of time, or if the study population can be expected to change over the course of treatment. Stratification should be considered when patients are recruited from many sites. Adaptive randomization schemes should be considered when investigators require that balance be achieved on multiple factors.

G.2. IMPLEMENTATION

Statistical quality control procedures are essential to ensure the validity of the data collection and computer entry methods. Key data variables should be run through checking programs to ensure, at a minimum, that the data are within the permissible range of possible values, that missing data are flagged, that patients meet inclusion and exclusion criteria, and that patients' data forms

are obtained in a timely fashion, as per protocol. Double entry of all keyed data is preferred. A random sample of data coded on data entry forms should be checked against original sources (e.g. forms from laboratories and/or the medical record).

Once study eligibility is validated, subjects are enrolled (and possibly stratified on baseline factors), assigned a study identification and thereby randomized to treatment following the predetermined randomization plan.

G.3. ANALYSIS

Generally, comparisons among treatment groups should be made as an 'intent-to-treat' analysis; that is, (1) patients should be counted in the treatment group to which they were randomly assigned, (2) the denominator for a treatment should be all patients assigned to that treatment, and (3) all events (whether believed to be related to the disease process under treatment or not) should be counted in the comparison(s) of primary interest.

An intent-to-treat analysis can lead to an underestimate of the true treatment effect, especially if compliance is low, there are many treatment crossovers, or the denominator includes many patients who could not be followed for the outcome of interest. Secondary analyses might then be carried out on completers (those staying on the program to study end), controlling for compliance levels. In general, analyses focused on an individual patient's longitudinal response, using composite (multidimensional) outcomes, should be encouraged. The outcome dimensions could include symptoms and/or structural measurements.

Some analytic methods used to compare treatments in trials are as follows:

G.3.1. Comparison of proportions

This method is valid provided that patients are subject to the same length of follow-up and the loss of follow-up is low, and occurs for the same reasons, across treatment groups. Statistical evaluation of the difference in proportions can be performed using Fisher's exact tests or chi-square test for larger samples. Examples include proportion who are 'pain free' and proportions experiencing serious adverse medical events. An example with respect to structure modifying drugs might include proportions developing joint space narrowing.

G.3.2. Lifetable analysis

This approach provides a means for dealing with varying duration of follow-up to achieve the primary endpoint and for dealing with cases where the primary endpoint does not occur by the end of the study ('censored data'). Statistical comparisons of lifetable rates are often performed using a 'log rank' test. Examples include: time until pain resolves, time until normalization of a laboratory parameter.

G.3.3. Comparison of means

This method is valid, subject to the same conditions required for comparing proportions (see above). Statistical evaluation of the difference in means can be performed using a two-sample t-test or the standard normal distribution for larger samples. Example: comparing average change in pain over the study period.

G.3.4. Descriptive methods

These are useful for assessing the baseline comparability of the treatment groups, and for secondary analyses assessing compliance issues. Descriptive statistics often include means, standard deviations, and percents of subjects in different strata (e.g. gender).

G.4. ADJUSTMENT PROCEDURES

To be valid, evaluation of treatment effects must be performed on treatment groups that are comparable with respect to their baseline characteristics. Statistical adjustment for one or more sources of variation is often performed by using regression models. Multiple linear regression models are used for quantitative outcomes, multiple logistic regression models are used for binary outcomes, and Cox proportional hazards models are used to adjust rates calculated from lifetables. These methods are especially useful if the randomization scheme failed or if randomization was not used in allocating patients to treatment groups.

G.5. INTERIM ANALYSIS

The concept of interim analysis is that patients assigned to the inferior treatment should be removed from it as soon as the choice is clear. These methods provide statistically valid P-values by accounting for the multiple looks of the outcome data during the study period. The scheme

for interim analyses should suit the particular trial. The procedure of O'Brien and Fleming is one statistically valid method for adjusting the *P*-value.

G.6. REPEATED MEASURES ANALYSES

These methods are useful for quantifying the trend and tempo of outcomes repeatedly assessed during the course of a trial and during the extended follow-up period. Statistical evaluation of the difference in summary statistics (e.g. trend, or slope) can be performed using the analysis of variance for repeated measures. Comparisons across treatment groups are valid provided that patients are followed for the same length of time and there is no differential loss to follow-up.

H. Summary

H.1. SYMPTOM MODIFYING DRUGS

The primary outcome variable is a specific aspect of joint pain, although a 'signal' symptom or some measure of function may also be studied. Trials of drugs with a rapid onset of effect can be as short as 1–4 weeks but may be as long as 12 weeks. Longer trials (up to 2 years) may be needed to evaluate longer-term toxicity, determine optimal long-term dosing regimens, or establish long-term benefit. Supplemental escape analgesia should be minimized, monitored and discontinued prior to evaluation of efficacy.

Some agents that provide symptom relief may not provide benefit until weeks after initiation of therapy. Under these circumstances, trials will vary from 3–12 months in length. If the agent is administered in courses, episodic readministration of the drug may be needed in long-term trials. Longer trials (up to 2 years) may be required to exclude toxicity or establish long-term benefit.

H.2. STRUCTURE MODIFYING DRUGS

These drugs are intended to prevent, retard, stabilize or reverse development of the morphologic changes of OA. Although this has been called 'chondroprotection', the term is misleading and should be avoided, because all structures of the joint are involved in OA, not articular cartilage alone. The benefits of disease modifying therapy may not be apparent until years after the onset of treatment. The selection of high-risk groups may shorten the time of investigation. Improvement in symptoms (i.e., joint pain) is not a requisite for the efficacy of a drug in this category. In these studies, it may be necessary to permit concomitant use of drugs for relief of symptoms (NSAIDs, analgesics). The confounding effects of glucocorticoids and NSAIDs in these trials is not yet understood and very restricted use of IA depocorticosteroids is recommended.

Demonstration of structure modification will require the use of direct measures of joint anatomy, such as radiography, particularly measurement of the radiographic joint space. As stated above, the plain radiograph is presently the most reproducible and readily available method for assessment of disease modification. Studies are needed to validate surrogate markers of disease activity, since they may help shorten Phase 2 structure modifying drug trials. As an alternative to radiography, some trials may utilize arthroscopy.

As we approach the beginning of the twenty-first century, concepts of clinical trials of OA drugs are changing. Methodology and techniques for the evaluation of new agents for OA have been refined dramatically over the last decade. We look forward to the future with excitement as we anticipate the development of new agents that may alter the symptoms and course of OA. The above recommendations are intended to help us ascertain which of these new agents are effective.

Acknowledgments

The Osteoarthritis Research Society would like to acknowledge the tireless and efficient work of our Executive Secretary, Evie Altman Orbach, especially in organizing the Task Force meetings and the Workshop. We wish to recognize the cooperation and participation of the International League Of Associations For Rheumatologists (ILAR) and the American College of Rheumatolgy. We wish to also recognize the participation and support of the European Group for the Respect of Ethics and Excellence in Science (GREES) and our industrial partners.

References

1. Lequesne M, Brandt K, Bellamy N, et al. Guidelines for testing slow acting drugs in osteoarthritis: J Rheumatol 1994;21Suppl 41:65–71.
2. Kuettner K, Goldberg V, editors. Osteoarthritic Disorders. Rosemont: American Academy of Orthopedic Surgeons, 1995.
3. Furst D. Guiding Principles for the development of drugs for osteoarthritis, 1995.
4. Group for the Respect of Ethics and Excellence in Science (GREES: osteoarthritis Section). Recommendations for the registration of drugs used in the treatment of osteoarthritis. Ann Rheum Dis 1996;55:552–7.

5. Bellamy N, Kirwan, J, Boers M, Brooks P, Strand V, Tugwell P, et al. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis: consensus development at OMERACT III. J Rheumatol 1996;In Press.
6. Altman RD, Hochberg MC: Degenerative joint disease, Clin Rheum Dis 1983;9:681–93.
7. Rataiu JS, Hochberg MC. Clinical trials: a guide to understanding methodology and interpreting results. Arthritis Rheum 1990;33:131–9.
8. FDA Document #77-3040. General considerations for the clinical evaluation of drugs, CDER Executive secretarial staff, HFD-8, FDA (Center for Drug Evaluation and Research, 5400 Fisher's lane, Rockville, MD 20857.
9. FDA Document: Guidelines for industry. The extent of population exposure to assess clinical safety: for drugs intended for long-term treatment of non-life-threatening conditions, ICH-EIA, March 1995; Federal Register March 1, 1995 (60FR11270).
10. Altman RD, Asch E, Bloch D, Bole G, Borenstein D, et al. Development of criteria for the classification and reporting of osteoarthritis: classification of osteoarthritis of the Knee. Arthritis Rheum 1986;29:1039–49.
11. ACR Subcommittee on Classification Criteria of Osteoarthritis—Altman RD, Chairman: The American College of Rheumatology Criteria for the Classification and Reporting of Osteoarthritis of the Hand. Arthritis Rheum 33:1990;1601–10.
12. ACR Subcommittee on Classification Criteria of Osteoarthritis—Altman RD, Chairman: The American College of Rheumatology Criteria for the Classification and Reporting of Osteoarthritis of the Hip. Arthritis Rheum 1991;34:505–14.
13. Kellgren JH, Lawrence JS. Radiological assessment of osteoarthritis. Ann Rheum Dis 1957;16:494–501.
14. Kellgren JH, The epidemiology of chronic rheumatism. Atlas of standard radiographs of arthritis, 2nd edn. Philadelphia, F. A. Davis Co. 1963:1–13.
15. Burnett S, Hart DJ, Cooper C, and Spector TD. A radiographic atlas of osteoarthritis. London. Springer Verlag 1994.
16. Kalman DA, Wigley FM, Scott WW, et al. New radiographic grading scales of osteoarthritis of the hand. Arthritis Rheum 1989;32:1584–91.
17. Altman RD, Hochberg M, Murphy WA, et al. Atlas of individual radiographic features in osteoarthritis. Osteoarthritis Cart 1995;3:3–70.
18. Spector TD, Hart DJ, Doyle DV. Indicence and progression of osteoarthritis in women with unilateral knee disease in the general population: the effect of obesity. Ann Rheum Dis 1994;53:565–8.
19. Neyret P, Donell ST, DeJour D, De Jour H. Partial meniscectomy and anterior cruciate ligament rupture in soccer players: a study with a minimum 20-yr follow-up. Am J Sports Med 1993;21:455–60.
20. Roos H, Lindberg H, Gardsell P, Lohmander LS, Wingstrand H. The prevalence of gonarthrosis and its relation to meniscectomy in former soccer players. Am J Sports Med. 1994;22:219–22.
21. Croft P, Cooper C, Wickham C, Coggon D. Defining osteoarthritis of the hip for epidemiologic studies. Am J Epidemiol 1990;132:514.
22. Hochberg MC, Lane NE, Pressman AR, Genant HK, Scott JC, Nevitt MC. The association of radiographic changes of osteoarthritis of the hand and hip in elderly women. J Rheumatol 1995;22:2291–4.
23. Sharif M, George E, Shepstone, L, et al. Serum hyaluronic acid level as a predictor of disease progression in osteoarthritis of the knee. Arthritis Rheum 1995;38:760–7.
24. Sharif M, Saxne T, Shepstone L, et al. Relationship between serum cartilage oligomeric matrix protein levels and disease progression in osteoarthritis of the knee joint. Br J Rheumatol 1995;34:306–10.
25. Bellamy N, Buchanan WW, Goldsmith CH, Campbell, Stitt LW. Validation of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988;15:1833–40.
26. Lequesne M, Meezy C, Samson M, et al. Indices of severity for osteoarthritis of the hip and knee. Scand J Rheumatol 1987;Suppl 65:85–9.
27. U.S. Food and Drug Administration Regulations 21 CFR Parts 50.20-50.27.
28. Haynes RB, Dantes R. Patient compliance and the conduct and interpretation of therapeutic trials. Controlled Clin Trials 1987;8:12–29.
29. Drummond MF. Principles of economic appraisal in health care. Oxford: Oxford University Press, 1980.
30. Ehrlich GE. Treatment decisions, side-effect liability and cost-effectiveness in osteoarthritis. Inflammopharmacology 1996;4:137–40.
31. Fries JF, Spitz P, Kraines RG, Holman JH. Measurement of patient outcome in arthritis. Arthritis Rheum 1980;23:137–45.
32. Meenan RF, Gurtman PM, Mason JH. Measuring health status in arthritis: the Arthritis Impact Measurement Scales. Arthritis Rheum 1980;23:146–54.
33. Mason JH, Anderson JJ, Meenan RF. Application of a health status model to osteoarthritis. Arthritis Care Res 1989;2:89–93.
34. Treves R, Maheu E, and Dreiser R-L. Therapeutic trials in digital osteoarthritis. A critical review. Revue Rheumatisme Engl edition 1994;626 Suppl 1:33S-41S.
35. Bellamy N, Haraoui B, Buckbinder R, et al. Development of a disease-specific health status measure for hand osteoarthritis clinical trials. 1. Assessment of the symptom dimensionality. Scand J Rheumatol 1996;Suppl 106:5.
36. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF36): I. Conceptual framework and item selection. Med Care 1992;30:473–83.
37. Bergner M, Bobbitt RA, Carter WB, Gilson BS. The Sickness Impact Profile: development and final revision of a health status measure. Med Care 1981;19:778–805.
38. Hunt S, McEwan P. The development of a subjective health indicator. Sociol Health Illness 1980;2:231–46.
39. Euroqol Group: Euroqol: A new facility for the measurement of health related quality of life. Health Policy 1990;16:199–208.
40. Torrance GW. Social preferences for health status. An empirical evaluation of the measurement techniques. Socio-economic Planning Sciences 1976;10:128–36.
41. Feeny D, Furlong W, Barr RD, et al. A comprehen-

sive multiattribute system for classifying the health status of survivors of childhood cancer. J Clin Oncol 1992;10:923–8.
42. Theiler R, Stucki G, Schutz R, Hofer H, Seifert B, Tyndall A, Michel BA. Parametric and non-parametric measures in the assessment of knee and hip osteoarthritis: interobserver reliability and correlation with radiology. Osteoarthritis Cart 1994;2:1–24.
43. Rejeski WJ, Ettinger WH Jr, Schumacker S, et al. Assessing performance related disability in patients with knee osteoarthritis. Osteoarthritis Cart 1995;3:157–68.
44. Felson DT, Anderson JJ, Boers M, Bombardier C, Furst D, Goldsmith C, et al. American College of Rheumatology Preliminary definition of improvement in rheumatoid arthritis. Arthritis Rheum 1995;38:727–35.
45. Hochberg MC. Quantitative radiography: osteoarthritis—analysis. Bailliere's Clin Rheum 1996;10:In Press.
46. Dougados M, Villers C, Amor B. Sensitivity to change of various roentgenological severity scoring systems for osteoarthritis of the hip. Rev Rhum 1995;[Engl Ed]62:169–84.
47. Lequesne M. Chondrometry quantitative evaluation of joint space width and rate of joint space loss in osteoarthritis of the hip. Rev Rhum 1995;[Engl Ed]62:155–8.
48. Buckland-Wright JC, Macfarlane DG, Williams SA, Ward RJ. Accuracy and precesion of joint space width measurements in standard and macroradiographs of osteoarthritic knees. Ann Rheum Dis 1995;54:872–80.
49. Ravaud P, Giraudeau B, Auleley GR, et al. Radiographic assessment of knee osteoarthritis: reproducibility and sensitivity to change. J Rheumatol 1996;23:1756–64.
50. Dieppe P, Altman R, Buckwalter J, et al. Standardization of methods used to assess the progression of osteoarthritis of hip and knee. In: New Horizons in Osteoarthritis, Kuettner K and Goldberg V, eds. Rosemont: American Academy of Orthopedic Surgeons 1995;481–96.
51. Buckland-Wright C. Protocols for precise radio-anatomical positioning of the tibiofemoral and patellofemoral compartments of the knee. Osteoarthritis Cart 1995;3(Suppl A):71–80.
52. Ravaud P, Auleley GR, Chastang C, et al. An experimental study of the influence of radiographic procedure and joint positioning on knee joint space width measurement. Br J Rheumatol 1996;35:761–6.
53. Mazzuca S, Brandt KD, Katz B. Is conventional radiography suitable for evaluation of a disease-modifying drug in patients with knee osteoarthritis. Osteoarthritis Cart. In press.
54. Buckland-Wright JC, Macfarlane DG. Radioanatomic assessment of therapeutic outcome in osteoarthritis. In: Kuettner KE, Goldberg VM, eds. Osteoarthritic Disorders, Rosement: American Academy of Orthopedic Surgeons, 1995, pp. 51–65.
55. Klashman D, Ike R, Moreland L, Skovron ML, Kalunian K. Validation of an osteoarthritis data report form for knee arthroscopy. Arthritis Rheum 1995;38(suppl 9):S178.
56. Ayral X, Dougados M, Listrat, Bonvarlet J-P, Simonnet J, Amor B. Arthroscopic evaluation of chondropathy in osteoarthritis of the knee. J Rheumatol 1996;23:698–706.
57. Poole AR. Immunochemical markers of joint inflammation, skeletal damage and repair; where are we now? Ann Rheum Dis 1994;53:3–5.
58. Lohmander LS. Articular cartilage and osteoarthrosis—the role of molecular markers to monitor breakdown, repair and disease. J Anat 1994;184:477–92.
59. Lark MW, Bayne EK, Lohmander LS. Aggrecan degradation in osteoarthritis and rheumatoid arthritis. Acta Orthop Scand 1995;66 Suppl 266:92–7.
60. Hully SB, Cummings SR. Designing Clinical Research. Baltimore, Williams and Willkins, 1988.
61. Meinert CL. Clinical Trials, New York: Oxford University Press. 1986.

Appendix I

Workshop participants

Code
[1] Organizing Chairperson
[2] Plenary Session Chairperson
[3] Speaker
[4] Chairperson—Breakout Session
[5] Co-chairperson—Breakout Session
[6] Scribe—Breakout Session

Roy Altman, Miami, FL, U.S.A. [2, 3]
Marliese Annofeld, Monza, Italy
Roberto Arinoviche, Santiago, Chile
Bernard Avouac, Creteil, France
Xavier Ayral, Paris, France [3, 5]
Nicholas Bellamy, London, ON, Canada [3, 4]
Daniel Bloch, Palo Alto, CA , U.S.A. [4]
Kenneth Brandt, Indianapolis, IN, U.S.A. [5]
Peter Brooks, Sydney, Australia [3]
J. C. Buckland-Wright, London, U.K. [3, 4]
Roland Chang, Chicago, IL, U.S.A.
Xavier Chevalier, Creteil, France
Cyrus Cooper, Southampton, U.K. [6]
Philipe Coste. Courbevoie, France
L. Crasborn, Brussels, Belgium
J. P. Devogelaer, Brussels, Belgium
Paul Dieppe, Bristol, U.K. [5]
Maxime Dougados, Paris, France [3]
Elliot Ehrich, Chatham, NJ, U.S.A.
George Ehrlich, Philadelphia, PA, U.S.A. [6]
Dominique Ethgen, Staines, U.K.
Walter Ettinger, Winston-Salem, NC, U.S.A. [6]
Yoshimasa Fujita, Tokyo, Japan
Daniel Furst, Seattle, WA, U.S.A. [3]
Peter Ghosh, Sydney, Australia
Klaus Peter Guenther, Ulm, Germany
George Bernard Guillou, Courbevoie, France
Laurie Hall, Cambridge, U.K.

Paul Haraoui, Montreal, PQ, Canada
Yves Henrotin, Liege, Belgium
Rosemarie Hirsch, Bethesda, MD, U.S.A.
Marc Hochberg, Baltimore, MD, U.S.A. [1, 2, 5]
Zeb Horowitz, Cincinnati, OH, U.S.A. [5]
Robert Ike, Ann Arbor, MI, U.S.A. [4]
T. Ishihara, Tokyo, Japan
Sergio Jimenez, Philadelphia, PA, U.S.A.
Kent Johnson, Bethesda, MD, U.S.A.
E. A. Jones, Amsterdam-Zuidoost, Netherlands
Joanne Jordan, Chapel Hill, NC, U.S.A.
Ken Kalunian, Los Angeles, CA, U.S.A.
Linda Katz, Rockville, MD, U.S.A. [4]
Seiichi Kobayashi, Ibaraki, Japan
Nancy Lane, San Francisco, CA, U.S.A.
Michael R. G. Leeming, Sandwich, U.K.
Richard Leff, West Haven, CT, U.S.A.
Arleette Lepot, Boulogne, France
Michel Lequesne, Paris, France [3]
Bruce Littman, Groton, CT, U.S.A.
Stefan Lohmander, Lund, Sweden [2, 4]
Leland Loose, Groton, CT, U.S.A.
Carlos Lozada, Miami, FL, U.S.A.
Ron Magolda, Wilmington, DE, U.S.A.
Emmanuel Maheu, Paris, France
Henry Mankin, Boston, MA, U.S.A.
Joe Markenson, New York, NY, U.S.A.
Joel Menkes, Paris, France [4]
Joan Meyer, Cincinnati, OH, U.S.A.
Christopher Mojcik, West Haven, CT, U.S.A.
Roland Moskowitz, Cleveland, OH, U.S.A.
Richard Newmark, Summit, NJ, U.S.A.
Pat O'Brien, Woburn, MA, U.S.A.
Michael O'Regan, Brindisi, Italy
Laurence Paolozzi, Paris, France
J.P. Pelletier, Montreal, PQ, Canada [2, 4]
Charles Peterfy, San Francisco, CA, U.S.A. [6]
Stanley Pillemer, Bethesda, MD, U.S.A.
Bernard Prigent, Kirkland, PQ, Canada
Ruth Raiss, Wiesbaden, Germany
Philippe Ravaud, Paris, France
J. Y. Reginster, Liege, Belgium [2]
Charles Rivest, Montreal, PQ, Canada
Harold Roos, Helsingborg, Sweden [6]
Teresa Rosario-Jansen, Cincinnati, OH, U.S.A. [5]
Lucio Rovati, Monza, Italy
Alan Rubinow, Jerusalem, Israel
Eric Sauvage, Courbevoie, France
Thomas Schnitzer, Chicago, IL, U.S.A. [5]
Benjamin Schwartz, St. Louis, MO, U.S.A. [6]
William Scott, Baltimore, MD, U.S.A. [5]
Umberto Serni, Florence, Italy
Vibeke Strand, San Francisco, CA, U.S.A.
Takashi Sugiyama, Tokyo, Japan
Eugene Thonar, Chicago, IL, U.S.A. [6]
J. Carter Thorne, Newmarket, ON, Canada
Stephen Trippel, Boston, MA, U.S.A. [6]
Y. Tsouderos, Courbevoie, France
Peter Tugwell, Ottowa, ON, Canada
Jenny Tyler, Cambridge, U.K.
Robert Vender, Wilmington, DE, U.S.A.
Gust Verbruggen, Ghent, Belgium [5]
Koju Watanabe, Tokyo, Japan
Michael Weisman, San Diego, CA, U.S.A.
Fred Wolfe, Wichita, KS, U.S.A.

Appendix II

Clinical assessment techniques

Nicholas Bellamy

Core set measures

The core set of outcome measures for OA clinical trials developed at OMERACT III, contain three clinical measures: pain, physical function and patient global assessment with imaging for studies of 1 year or longer [1].

PAIN

Pain is usually measured on a rating scale (Likert or VAS) which grades perceived pain severity in one or several situations (e.g., nocturnal, stair climbing, walking, rest, global) [2]. The pain subscale of the WOMAC OA Index has been validated for use in patients with hip and/or knee OA [3–6]. WOMAC is available in both Likert and 10 cm VA scaled formats and in a large number of alternate language translations. Although not recommended for use as a distinct pain scale, the AFI have been validated for use in hip and knee studies where the goal is to provide a weighted clinical severity score in which scores for pain/discomfort, stiffness, maximum distance walked and activities of daily living are summated into a single value [7, 8]. A similar approach can be used with the WOMAC in situations where a composite score (based on pain, stiffness and physical function) is required, using weights derived from the Patient Assessment of the Relative Importance of Symptoms (PARIS) Sectogram [6]. The Health Assessment Questionnaire (HAQ) [9] pain scale or the AIMS [10] or AIMS2 [11] may be of limited value for studies focusing on a single joint, because they are appropriated for studies measuring pain severity in both the upper and lower extremities. Options for OA hand studies are limited but early experience with the pain subscale of an instrument termed the Australian/Canadian (AUSCAN) Osteoarthritis Hand Index has been favorable [12].

PHYSICAL FUNCTION

Physical function/disability is usually measured on a rating scale (Likert, VAS) which grades the perceived severity or degree of disability in one or more activities of daily living (e.g., stair climbing, walking, etc) [13]. The physical function subscale of the WOMAC index has been validated for use in patients with hip and/or knee OA [3]. The index is available in both Likert and 10 cm VAS scaled formats and in a large number of alternate language translations. Although not recommended as a distinct physical function scale, the AFI have been validated for use in hip and knee studies, where the goal is to provide a weighted clinical severity score in which scores for pain/discomfort, stiffness, maximum distance walked and activities of daily living are summated into a single score [7, 8]. A similar approach can be used with the WOMAC by weighting and aggregating the pain, stiffness and physical function subscales using PARIS Sectogram weights [6]. An AFI developed by Dreiser and colleagues contains 10 questions directed at functional disability in the hand [14]. Early experience with the function subscale of the AUSCAN Osteoarthritis Hand Index has been favorable [12]. In studies measuring physical disability in both the upper and lower extremities, the physical function subscale of the HAQ [9] or AIMS [10] (or AIMS2) [11] instruments is appropriate.

PATIENT GLOBAL ASSESSMENT

The patient's perception of the clinical severity of their OA is usually assessed by a direct question, e.g., 'Considering all the ways your OA affects you, how would you rate your condition today?' Suitable response scales could include the following: Likert—very poor, poor, fair, good, very good; or 10 cm horizontal VAS anchored to very poor (left hand end) and very good (right hand end). Alternatively, or in addition, at the end of the study a change in score could be derived using a similar question, e.g., 'Considering all the ways your OA has affected you, how do you feel now compared with the beginning of the study?' Responses could be made on a Likert scale, e.g., 'much better', 'better', 'no change', 'worse', 'much worse'. There is currently no standard question and no standard response format [15]. It should be noted that depending on the research hypothesis, there are several ways of phrasing the global question, e.g., musculoskeletal condition, OA in the study knee, overall health, etc. Investigators should be guided by questions and response formats that have been used successfully in past studies or should develop and validate new standardized questions.

Non-core set measures

HEALTH-RELATED QUALITY OF LIFE (HRQOL) AND UTILITY (UT) MEASURES

HRQOL and/or UT measures are increasingly being considered as very important components of the measurement battery for studies of 6 months or longer. They not only allow measurement of the patient's quality of life or the utility of their health state, but also facilitate pharmacoeconomic and cross-disease comparisons of outcome. There is relatively little experience to date with these instruments (e.g., SF-36 [16], EuroQol [17], Nottingham Health Profile [18], Health Utilities Index [19], Standard Gamble [20], Time Trade Off [21], Category Scaling [22]) in OA trials. However, because of their potential importance, use of HRQOL and/or UT measures is highly recommended in Phase 3 trials of 6 months or longer. It is expected that there will be improvement in our knowledge of the performance of one or more of these instruments, their role in Phase 3 clinical trials and the relative impact that interventions have on different measures will evolve over the next few years. Comparing different measures in the same trial would be particularly useful.

PHYSICIAN GLOBAL ASSESSMENT

The physician's perception of his or her patient's OA can be based on a number of different variables, e.g., symptoms, signs, imaging, and, possibly, in the future, biologic markers. It is important to specify in the question or in accompanying instructions which variables should be considered in making the assessment. Usually, this will be based on symptoms, and since the clinical encounter will likely be quite brief, the question should be phrased with respect to the day of assessment, e.g., 'Considering all the ways OA affects your patient, how would you rate his or her condition today?' Suitable response scales could include the following: Likert—very poor, poor, fair, good, very good; or 10 cm horizontal VAS anchored to very poor (left hand end) and very good (right hand end). Alternatively, or, in addition, at end of study a change score can be derived using a similar question, e.g., 'Considering all the ways OA has affected your patient, how do you rate their condition now compared with the

beginning of the study?' Responses could be made on a Likert scale, e.g., much better, better, no change, worse, much. There is currently no standard question and no standard response format [15]. Investigators should be guided by questions and response formats that have been used successfully in past studies or should develop and validate new standardized questions. It should be noted that depending on the research hypothesis, there are several ways of phrasing the global question, e.g., musculoskeletal condition, OA in study knee, overall health, etc.

PERFORMANCE-BASED MEASURES

Many performance based measures are available, some of which are of demonstrated reliability, validity and responsiveness [15, 23]. Although providing numerical estimates of performance, the clinical consequence to individual patients of any change for the better or worse on such measures lacks clarity. As a consequence, while sometimes useful in certain types of studies, they are not included in the core set. Measures that have been employed include: walking distance, walk time, grip strength [23]. It is important with these measures to use standard techniques and to train assessors to acceptable levels of inter-observer reliability [15].

EXAMINATION BASED MEASURES

The clinical examination provides an opportunity to detect swelling (bony, soft tissue, effusion), crepitus, heat, range of movement, deformity, ligamentous laxity, range of movement (goniometer, plurimeter, intermalleolar straddle, intercondylar distance, heel to buttock test) [15, 23]. These assessments require standard methods applied by trained assessors [15]. In general, as with performance-based measures, changes for the better or worse occurring on these examination-based measures lack defined levels of clinical importance to individual patients. They may be useful in some types of study but are not in the core set.

MISCELLANEOUS

'Stiffness' is a sense of resistance or decreased ease during active movement of the joint. Some, but not all, patients have difficulty differentiating between pain and stiffness. When stiffness is measured in clinical trials, it is preferable to use the WOMAC [3–6] or AFI [7, 8] (depending on whether a segregated or aggregated stiffness score is required) for hip and knee studies. The assessment of stiffness may be useful in some types of studies but has not been validated as an outcome in OA and is not in the core set.

'Inflammation' has not been extensively studied in OA clinical trials. As a result, the validity, reliability and responsiveness of inflammatory-based measures remain in doubt. They may be useful in some types of study but are not validated in OA and are not in the core set.

'Number of "Flares"' and the occurrence of disease 'flares' in OA lacks precise definition and as a result is difficult to reliably identify. This variable has not been included in the core set.

'Time to surgery' is influenced by a large number of factors, independent of the study intervention, e.g., the dynamics of scheduling operating time. Although this variable may be useful in some studies it has not been validated as an outcome in OA and is not included in the core set.

'Analgesic consumption' is an important source of cointervention; however, the precision with which analgesic consumption can be monitored, particularly in long-term studies, is suboptimal.

RESPONSE CRITERIA

Response criteria may apply to groups of patients or individuals. The definition of a minimum clinically important difference between two groups of patients exposed to different interventions, depends on a number of factors relating to patient characteristics, disease features, the nature of the interventions, and the primary outcome measure selected. It is difficult to determine estimates of minimum clinically important differences for OA clinical studies [15]. At the present time there are no standard criteria for defining the success, or failure, of treatment in individual OA patients in a clinical trial.

Summary

The core set of outcome measures for OA clinical trials requires measurement of pain, physical function, patient global assessment, and imaging procedures for studies of 1 year or longer. Depending on the research hypothesis, one of several existing validated measures can be selected for evaluating change in each of the four domains.

References

1. Bellamy N, Kirwan J, Boers M, et al. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip and

hand osteoarthritis. Consensus development at OMERACT III. J Rheumatol 1996;In press.
2. Bellamy N. Pain measurement. In: Musculoskeletal Clinical Metrology. Dordrecht: Kluwer Academic Publishers, 1993:65–76.
3. Bellamy, N. Osteoarthritis—An evaluative index for clinical trials. MSc Thesis, McMaster University, Hamilton, Canada. 1982.
4. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient-relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988;15:1833–40.
5. Bellamy N, Campbell J, Buchanan WW, Goldsmith CH, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically-important patient-relevant outcomes following total hip or knee arthroplasty in osteoarthritis. J Orthopaed Rheumatol 1988;1:95–108.
6. Bellamy N. WOMAC Osteoarthritis Index—A User's Guide. London: London Health Science Centre, 1996.
7. Lequesne M, Mery C, Samson M, Gerard P. Indexes of severity for osteoarthritis of the hip and knee: validation-value in comparison with other assessment tests. Scandinavian J Rheumatol 1987;Suppl 65:85–9.
8. Lequesne M. Indices of severity and disease activity for osteoarthritis. Semin Arthritis Rheum 1991;20:48–64.
9. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. Arthritis Rheum 1980;23:137–45.
10. Meenan RF, Gurtman PM, Mason JH. Measuring health status in arthritis: the Arthritis Impact Measurement Scales. Arthritis Rheum 1980;23:146–52.
11. Meenan RF, Mason JH, Anderson JJ, Guccione AA, Kazis LE. AIMS2. The content and properties of a revised and expanded Arthritis Impact Measurement Scales Health Status Questionnaire. Arthritis Rheum 1992;35:1–10.
12. Bellamy N, Haraoui B, Buchbinder R, et al. Development of a disease-specific health status measure for hand osteoarthritis clinical trials. I. Assessment of symptom dimensionality. [Abstract] Scand J Rheumatol 1996;Suppl 106:5.
13. Bellamy N. Health status instruments and functional indices. Musculoskeletal Clinical Metrology. Dordrecht: Kluwer Academic Publishers, 1993:77–101.
14. Dreiser RL, Maheu E, Guillou GB, Caspard J, Grouin JM. Validation of an algofunctional index for osteoarthritis of the hand. Revue du rhumatisme 1995;(English Edn)Suppl 1:43S–53S.
15. Bellamy N. Standardized procedures for outcome measurement and parameters for calculating sample size for antirheumatic-drug studies. In Musculoskeletal Clinical Metrology. Dordrecht: Kluwer Academic Publishers, 1993:193–251.
16. Ware JE, Sherbourne CD. The MOS 36-item Short-Form Health Status Survey (SF-36): I. Conceptual framework and item selection. Medical Care 1992;30:473–83.
17. Economic and Health Outcome Research Group: Jobanputra P, Hunter M, et al. Validity of Euroqol—a generic health status instrument in patients with rheumatoid arthritis. Br J Rheumatol 1994;33:655–62.
18. The European Group for Quality of Life and Health Measurement. European Guide to the Nottingham Health Profile, 1992.
19. Feeny D, Furlong W, Barr RD, et al. A comprehensive multiattribute system for classifying the health status of survivors of childhood cancer. J Clin Oncol 1992;10:923–8.
20. Von Neumann J, Morgenstern O. Theory of games and economic behaviour. Princeton NJ: Princeton University Press; 1953.
21. Torrance GW, Thomas WH, Sackett DL. A utility maximization model for evaluation of health care programs. Health Services Research. 1972;7:118–33.
22. Torrance GW. Social preferences for health states: an empirical evaluation of three measurement techniques. Socioeconomic Planning Sciences. 1976:10(3):128–36.
23. Bellamy N. Mechanical and electromechanical devices. In Musculoskeletal Clinical Metrology. Dordrecht: Kluwer Academic Publishers, 1993:117–34.

Appendix III

Radiographic imaging techniques

J. Christopher Buckland-Wright,
William W. Scott Jr, Charles Peterfy

The reproducibility of the radiographic technique is dependent on control of a number of technical issues. The discussion below presents a few methods that attempt to standardize many of the relevant techniques. Other methodologies exist. Such standardization is essential in order to reliably assess sequential changes in joint anatomy. The most consistent results will be obtained by carefully adhering to standardized radiographic procedures, based on published, validated data. Quality control of personnel and procedures is essential for multicenter or comparative studies. The methods described below require no special facilities other than fluoroscopy.

Hip joint

PATIENT POSITION

Anteroposterior radiographs are obtained with the patient standing. Weight bearing compresses the joint space to its most narrow configuration [1, 2]. The feet are positioned in internal rotation with the toes subtending an angle of $15 \pm 5°$ [3]. A foot map, used to facilitate repositioning at successive visits may improve measurement reproducibility. However, a foot map alone does not assure identical repositioning as the body can

torque about the knee. Reproducibility requires multipoint control.

X-RAY BEAM ALIGNMENT

With a focus to film distance of 100 cm, the X-ray beam must be horizontal and perpendicular to the film. When the X-ray beam is centered on the superior aspect of the symphysis pubis to radiograph both hips together. There is less accuracy and precision in the joint space width measurement than when the central ray of the X-ray beam is aligned with the center of each femoral head [3–5].

RADIOGRAPHIC MAGNIFICATION

In view of the variable distance between hip joint and film among individuals, variable radiographic magnification can occur. A metal sphere of known size (10 mm), mounted in a semi-radiolucent material and taped to the skin over the greater trochanter can be used to correct for radiographic magnification at the joint. This is needed only if significant weight change has occurred between visits to alter this distance. An increase in the number of study patients may be needed without correction for radiographic magnification.

Knee joint: tibio-femoral compartment

STANDING FULLY EXTENDED VIEW [6–9]

Patient position

Separate anteroposterior radiographs of each knee are obtained with the patient standing and the weight equally distributed to both feet. The knee must be in full extension, with the back of the knee as near as possible to the vertical cassette. With the aid of fluoroscopy, the lower limb is rotated so that the tibial spines appear centrally placed relative to the femoral notch. A foot map may be used to facilitate repositioning at successive visits.

X-ray beam alignment

The central ray of the X-ray beam is centered on the joint space and inclined downward to ensure that the medial tibial plateau is parallel to the X-ray beam.

Correction for radiographic magnification

It is only necessary to correct for radiographic magnification if the distance between the back of the knee and the vertical cassette is altered in subsequent examinations (see above).

STANDING PARTIALLY FLEXED VIEW [3, 10, 11]

Patient position

Separate anteroposterior radiographs of each knee are obtained with the patient standing. Each knee is flexed until the tibial plateau is horizontal relative to the floor, and therefore parallel to the central X-ray beam which is oriented perpendicular to the X-ray film. The degree of flexion varies among individuals due to differences in the angle of inclination of the tibial plateau. The precise inclination is obtained with the aid of fluoroscopy. With the heel fixed, the foot is internally or externally rotated until the tibial spines appear centrally placed relative to the femoral notch. A foot map may be used to facilitate joint repositioning at successive visits; patients are provided with hand supports to ensure their stability.

X-ray beam alignment

With a focus to film distance of 100 cm, the X-ray beam, must be horizontal to the floor, perpendicular to the film, and aligned with the center of the joint.

Radiographic magnification

Correction for the effect of radiographic magnification includes a metal sphere of known size (5 mm) taped above the head of the fibula. The dimension of this ball is used to determine the degree of radiographic magnification at the joint. This is only needed if there is variation in knee to film distance between visits.

Published studies may guide the calculation of numbers needed for a structure modifying drug trial [12, 13].

Wrist and hand joints

PATIENT POSITION

Dorsopalmar radiographs of the wrist and hand are obtained with the fingers held together and in line with the axis of the wrist and forearm, since spreading the fingers may alter joint alignment and lead to an incorrect assessment of joint space

loss. A hand map may be used to facilitate precise repositioning at successive visits.

X-RAY BEAM ALIGNMENT

The tube is positioned at a focus to film distance of 100 cm. The central ray of the X-ray beam is centered vertically at the head of the third metacarpal bone.

Landmarks for measurement

JOINT SPACE WIDTH OR INTERBONE DISTANCE

Joint space narrowing correlates with cartilage thickness in OA [13]. Because cartilage loss in OA is not uniform across the joint [14], minimum joint space width is the appropriate measurement [15].

References

1. Lequesne M. Quantitative measurement of joint space during progression of osteoarthritis: 'chondrometry'. In: Kuettner KE, Goldberg V, Eds. Osteoarthritic disorders. Rosemont: American Academy of Orthopedic Surgeons 1995;427–44.
2. Conrozier T, Tron AM, Mathieu P. Vignon E. Quantitative assessment of radiographic normal and osteoarthritic hip joint space. Osteoarthritis Cart 1995;3Suppl A:81–7.
3. Buckland-Wright JC. Quantitation of radiographic change. In: Brandt KD, Lohmander S, Doherty M, Eds Textbook of Osteoarthritis. Oxford: University Press 1997:In press.
4. Fife R, Brandt K, Braunstein E, et al. Relationship between arthroscopic evidence of cartilage damage and radiographic evidence of joint space narrowing in early osteoarthritis of the knee. Arthritis Rheum 1991;34:377–82.
5. Lynch JA, Buckland-Wright JC, Macfarlane DG. Precision of joint space width measurement in knee osteoanhritis from digital image analysis of high definition macroradiographs. Osteoarthritis Cart 1993;1:209–18.
6. Dieppe PA. Recommended methodology for assessing the progression of osteoarthritis of the hip and knee joint. Osteoarthritis Cart 1995;3:73–7.
7. Ravaud P, Auleley G, Chastang C, et al. Knee joint space width measurement: an experimental study of the influence of radiographic procedure and joint positioning. Br J Rheumatol 1996;35:761 6.
8. Ravaud P, Giraudeau B, Auleley G, et al. Radiographic assessment of knee osteoarthritis: reproducibility and sensitivity to change. J Rheumatol 1996;23:1756–64.
9. Ravaud P, Chastang C, Auleley G, et al. Assessment of joint space width in patients with osteoarthritis of the knee: a comparison of four measuring instruments. J Rheumatol 1996;23:1–15.
10. Buckland-Wright JC. Protocols for precise radioanatomical position of the tibiofemoral and patellofemoral compartment of the knee. Osteoarthritis Cart 1995;3 Suppl A:71–80.
11. Buckland-Wright JC, Macfarlane DO, Williams SA, Ward RJ. Accuracy and precision of joint space width measurement in standard and macroradiographs of osteoarthritic knees. Ann Rheum Dis 1995;54:872–80.
12. Mazzuca S, Brandt KD, Katz B. Is conventional radiography suitable for evaluation of a disease-modifying drug in patients with knee osteoarthritis. Osteoarthritis Cart. In press.
13. Buckland-Wright JC, Macfarlane DG, Lynch JA, Jasani MK, Bradshaw CR. Joint space width measures cartilage thickness in osteoarthritis of the knee: high resolution plain film and double contrast macroradiographic investigation. Ann Rheum Dis 1995;54:263–8.
14. Thomas RM, Resnick D, Alazraki NP, Deel D, Greenffeld R. Compartmental evaluation of osteoarthritis of the knee. A comparative study of available diagnostic modalities. Radiology 1975;116:585–94.
15. Lequesne M, Brandt K, Bellamy N, Moskowitz R. Menkes CJ, Pelletier J-P. Guidelines for testing slow acting drugs in OA. Proceedings Vth Joint WHO and ILAR Task Force Meeting. J Rheumatol 1994;21 Suppl 41:65–71.

Appendix IV

Magnetic resonance imaging

Charles Peterfy

MRI is a relatively new imaging technique, but its utility for evaluating structural derangements of diarthrodial joints, such as meniscal tears, cruciate ligament ruptures and bone injuries is already well-established in clinical practice. Recent techniques show promise for serial quantification of the volume, thickness, geometry and composition of articular cartilage [1]. These techniques are so new, however, that only a few have been validated cross-sectionally and none has been assessed longitudinally.

Possible uses of MRI

MRI is uniquely suited for monitoring structural changes in OA, for it is capable of directly examining all components of the joint simultaneously. In addition to delineating anatomy, however, MRI shows promise for quantification of compositional and functional parameters of articular tissues.

MEASURING CARTILAGE MORPHOLOGY

Fat-suppressed, T1-weighted three-dimensional (3D) gradient echo imaging can delineate articular

cartilage morphology in the knee [2–6]. and fingers [7, 8]. In a recent study of 48 knees [4, 5], this technique demonstrated a sensitivity of 86% and specificity of 97% for identifying cartilage defects which were visible on arthroscopy. The surface topography of individual cartilage plates as well as contact-areas between opposing articular surfaces can be mapped [9]. Accurate measurement of cartilage thickness requires spatial resolution better than 10% of that thickness (e.g., 200 μm in-plane resolution for a 2-mm thick cartilage). This is possible with conventional MRI, but generally beyond what is performed during routine clinical imaging. Considerably less resolution is required to quantify cartilage volume in the knee [3] or the metacarpophalangeal [7]. Validation of the longitudinal reproducibility of cartilage volume measurement and its sensitivity to volume changes will not be available for several years.

MEASURING CARTILAGE QUALITY

MRI may be able to probe the composition of articular cartilage. Areas of matrix loss and increased water content in the cartilage may cause focal signal intensity alterations and it may be possible to map the fractional water content of normal and abnormal cartilage [10]. This technique, however, must await further optimization and validation. Other parameters of articular cartilage integrity, such as water diffusivity [11, 12], proteoglycan content [13, 14], collagen content and organization [15, 16] and compressive stiffness, may be measurable in the future. Should these advancements occur, MRI may replace radiography as the standard imaging method.

EVALUATING OTHER ARTICULAR TISSUES IN OA

MRI also provides information about the severity of synovial inflammation, the integrity of IA ligaments and menisci, the status of periarticular muscles and tendons, the presence of subarticular bone marrow edema, and the morphology of the articular bones (including the size, number and location of osteophytes and subchondral cysts) [17, 18].

References

1. Peterfy CG, Genant HK. Emerging applications of magnetic resonance imaging for evaluating the articular cartilage. Radiol Clin N Am 1996;34:195–213.
2. Peterfy CG, Majumdar S, Lang P, van Dijke CF, Sack K, et al. MR imaging of the arthritic knee: improved discrimination of cartilage, synovium and effusion with pulsed saturation transfer and fat-suppressed T1-weighted sequences. Radiology 1994;191:413–9.
3. Peterfy CG, van Dijke CF, Janzen DL, Gluer C, Namba R, Majumdar S, et al. Quantification of articular cartilage in the knee by pulsed saturation transfer and fat-suppressed MRI: optimization and validation. Radiology 1994;192:485–91.
4. Recht MP, Piraino DW, Paletta GA, Schils JP, Belhobek GH. Accuracy of fat-suppressed three-dimensional spoiled gradient-echo FLASH MR imaging in the detection of patellofemoral articular cartilage abnormalities. Radiology 1996;198:209–12.
5. Disler DG, McCauley TR, Kelman CG, Fuchs MC, Ratner LM, Wirth CR, et al. Fat-suppressed three-dimensional spoiled gradient-echo MR imaging of hyaline cartilage defects in the knee: comparison with standard MR imaging and arthroscopy. AJR 1996;167:127–32.
6. Recht MP, Kramer J, Marcelis S, Pathria M, Trudell D, Haghighi P, et al. Abnormalities of articular cartilage in the knee: analysis of available MR techniques. Radiology 1993;187:473–8.
7. Peterfy CG, van Dijke CF, Lu Y, Nguyen A, Connick T, Kneeland B, et al. Quantification of articular cartilage in the metacarpophalangeal joints of the hand: accuracy and precision of 3D MR imaging. AJR 1995;165:371–5.
8. Robson M, Hodgson R, Herrod N, Tyler J, Hall L. A combined analysis and magnetic resonance imaging technique for computerised automatic measurement of cartilage thickness in the distal interphalangeal joint. Magnetic Resonance Imag 1995;13:709–18.
9. Ateshian GA, Cohen ZA, Kwak SD, Wang V, Kelkar R, Raimondo R, et al. Determination of in situ contact areas in diarthroidial joints by MRI. International Mechanical Engineering Congress and Exposition. San Francisco, CA, November 1995.
10. Selby K, Peterfy CG, Cohen ZA, Ateshian GA, Mow VC, Roos M, et al. In vivo MR quantification of articular cartilage water content: a potential early indicator of osteoarthritis. Book of Abstracts: Society of Magnetic Resonance 1995. Berkeley: Society of Magnetic Resonance, 1995:204.
11. Burstein D, Gray ML, Hartman AL, Gipe R, Foy BD. Diffusion of small solutes in cartilage as measured by nuclear magnetic resonance (NMR) spectroscopy and imaging. J Orthop Res 1993;11:465–78.
12. Xia Y, Farquhar T, Burton-Wuster N, Ray E, Jelinski LW. Diffusion and relaxation mapping of cartilage-bone plugs and excised disks using microscopic magnetic resonance imaging. Magn Reson Med 1994;31:273–82.
13. Kusaka Y, Grunder W, Rumpel H, Dannhauer K-H, Gersone K. MR microimaging of articular cartilage and contrast enhancement by manganese ions. Magn Reson Med 1992;24:137–48.
14. Eng J, Ceckler T, Balaban RS. Quantitative 1 H magnetization transfer imaging in vivo. Magn Reson Med 1991;17:304–14.
15. Rubenstein JD, Kim JK, Morava-Protzner I, Stanchev PL, Henkelamn RM. Effects of collagen Orientation on MR imaging characteristics of bovine cartilage. Radiology 1993;188:219–26.

16. Peterfy CG, Genant HK. Applications of magnetic resonance imaging in arthritis. In W. Koopman, J., ed. Arthritis and Allied Conditions. A Textbook of Rheumatology. Lea and Sebiger, 1996:In press.
17. Resnick D. Internal derangements of joints. In: D. Resnick, ed. Diagnosis of Bone and Joint Disorders. Philadelphia: W.B. Saunders, 1995:3063–3069.
18. Bell RA. Economics of MRI technology. JMRI 1996;6:10-2519. Peterfy CG, Roberts T, Genant HK. Dedicated extremity MRI: an emerging technology. In J.B. Kneeland, ed. Radiol Clin N Am. Philadelphia: W.B. Saunders, 1996:In press.

Appendix V

Methods for collection and storage of body fluid samples

Stefan Lohmander

Most of the published studies on markers have focused on analyzing cartilage-derived products; however, markers of the metabolism of other joint tissues, such as meniscus, synovium and bone should also be considered. In addition, markers of genetic susceptibility, and cellular activity or other processes that might be relevant to the pathogenesis of OA should also be given consideration. The following represents an update and summary of the previously published guidelines for sample collection and storage [1].

Sample collection

Three biological fluids are potential sources for markers in OA studies: urine, blood and synovial fluid. Guidelines for collecting these samples should minimize manipulations at the site of collection. While this probably does not present difficulties with urine and blood, some special problems are noted below for synovial fluids. In general, samples should be processed so that they may be frozen at the collection site in small screw-cap tubes designed for storage. Additional manipulations, involving dilution, aliquoting, storage and shipping of collected specimens to laboratories performing the marker assays would be best accomplished by a referral center with appropriate facilities and trained personnel. Due to the possibility of circadian variations in marker levels, care should be taken to collect samples at the same time of the day in longitudinal studies.

URINE

Specimens should be obtained as the second void in the morning; spot sampling would be acceptable; however, time of collection must be recorded. Specimens should be chilled to 4°C and clarified in a clinical centrifuge within 4 h. Approximately 25 ml should be aliquoted into a 50 ml polypropylene tube with a screw cap. The specimen should be clearly labeled and stored frozen (see below).

BLOOD

Approximately 25 ml blood should be taken from the antecubital vein after fasting, and collected in either plain, heparin or EDTA tubes. The choice of tubes is dictated by the effect either may have on the marker assays eventually chosen. For example, some heparin samples may contain interfering substances in assays for carbohydrate epitopes on chondroitin sulfate or keratan sulfate. The samples should be kept at 4°C until plasma (or serum) can be prepared by centrifugation in a clinical centrifuge, preferably within 4 h. The clarified plasma (or serum) should be aliquoted into 'Eppendorf-type' tubes (1 ml per tube). The tubes should be clearly labeled and stored frozen (see below). Although the collection of serum may be simpler, it was argued that plasma samples could be preferred for some marker assays, and may also be a source of DNA for analysis of genetic susceptibility. If analysis of genetic material is planned as a specific target, preparation and frozen storage of buffy coat is recommended.

SYNOVIAL FLUID

Synovial fluid should, if at all possible, be collected undiluted, i.e. without lavage. In cases without joint swelling and exudate, synovial fluid volumes will be small. Up to 10 aliquots of 1 ml should be distributed into 1.5–2.0 ml 'Eppendorf-type' tubes. Any remaining larger volumes can be stored in larger size aliquots. The tubes should contain EDTA in appropriate amounts to prevent fibrin clot formation. The tubes should be suitable for centrifugation in a higher speed centrifuge, such as a microfuge. The higher speeds are required to remove cells and debris from the samples which are frequently very viscous. Samples should be kept at 4°C and centrifuged within 4 h. Clarified supernatants should be transferred into appropriate sized (2 or 20 ml) polypropylene tubes with screw caps. The specimens should be clearly labeled and stored frozen, preferably at −70°C, prior to shipment to a referral center. A recommendation for sample centered information that should always be available, to complement the core clinical data, is included (Table 1).

Table I.
Sample centered data to be collected with all specimens

Urine
First or subsequent a.m. void, or other time of 'spot' sample

Serum
Site of venepuncture
(if not antecubital vein, where)

Synovial fluid
Total volume withdrawn
Lavage used (yes/no)—if yes, volume

For all
Date and time sample was taken
Have guidelines for handling and storage been adhered to
(if no, provide details)
At what temperature have samples been stored

Referral collection centers

The collection center should have defined protocols for thawing, diluting, aliquoting, coding (consider bar coding), freezing and storing the specimens received. Accompanying clinical and chemical data, required by the clinical protocol and any accessory information, would be encoded into a data bank system. This center would also distribute appropriate sample sets to laboratories conducting the marker assays. Results of the assays would be sent to this center and entered into the data bank. For synovial fluid samples, volumes will often be small, and a minimum set of aliquots could be prepared based upon a 1 ml volume. A measured volume, e.g. 1 ml, should be diluted with four volumes of physiological saline supplemented with either EDTA or heparin depending upon the original choice. Aliquots of 250 µl should then be distributed into small volume, coded polypropylene tubes with O-rings and screw caps. This dilution and sampling protocol will yield around 20 identical samples for each original synovial fluid sample. For synovial fluid samples with larger volumes, we recommend preparing at least one such set of 20 identical standards, and then storing the remainder of the 1:4 diluted samples in larger aliquots (5 or 10 ml) in appropriate tubes for long-term storage. All storage should be at –70°C.

Freezing and shipping

Freezing and thawing of samples should be minimized, as some markers may loose antigenicity in the process. Storage at –70°C is preferable if such freezer capabilities are available at the site of sample collection. In this case, the samples can be stored for long periods of time. If this option is not available, samples can be stored at –20°C in a freezer which does not have an automatic defrost cycle. Samples collected during a week should then be sent on dry ice to a referral collection center.

References

1. Dieppe P, Altman R, Buckwalter J, Felson D, Hascall V, Kuettner K, et al. Standardization of methods used to assess the progression of osteoarthritis of hip and knee. In: Kuettner K, Goldberg V. Eds. Osteoarthritic Disorders. Rosemont: American Academy of Orthopedic Surgeons 1995;481–96.

PAIN

Pain 106 (2003) 337–345

www.elsevier.com/locate/pain

Core outcome domains for chronic pain clinical trials: IMMPACT recommendations

Dennis C. Turk[a,*], Robert H. Dworkin[b], Robert R. Allen[c], Nicholas Bellamy[d], Nancy Brandenburg[e], Daniel B. Carr[f], Charles Cleeland[g], Raymond Dionne[h], John T. Farrar[i], Bradley S. Galer[j], David J. Hewitt[k], Alejandro R. Jadad[l], Nathaniel P. Katz[m], Lynn D. Kramer[n], Donald C. Manning[o], Cynthia G. McCormick[p], Michael P. McDermott[b], Patrick McGrath[q], Steve Quessy[r], Bob A. Rappaport[s], James P. Robinson[t], Mike A. Royal[u], Lee Simon[s], Joseph W. Stauffer[v], Wendy Stein[w], Jane Tollett[x], James Witter[s]

[a]*Department of Anesthesiology, University of Washington, Seattle, WA 98195, USA*
[b]*University of Rochester, Rochester, NY, USA*
[c]*AstraZeneca, Wilmington, DE, USA*
[d]*University of Queensland, Australia*
[e]*Pfizer Pharmaceutical Group New York, NY, USA*
[f]*Department of Anesthesiology, Tufts University, Medford, MA, USA*
[g]*MD Anderson Cancer Center, Houston, TX, USA*
[h]*National Institute of Dental and Craniofacial Research, Bethesda, MA, USA*
[i]*Center for Clinical Epidemiology & Biostatistics, University of Pennsylvania, Philadelphia, PA, USA*
[j]*Endo Pharmaceuticals Inc., Chadds Ford, PA, USA*
[k]*Ortho-McNeil Pharmaceutical Inc., Raritan, NJ, USA*
[l]*University Health Network and University of Toronto, Toronto, ON, Canada*
[m]*Harvard University, Cambridge, MA, USA*
[n]*Purdue Pharma, Stamford, CT, USA*
[o]*Novartis Pharmaceuticals Corporation, East Hanover, NJ, USA*
[p]*National Institutes of Health, Bethesda, MD, USA*
[q]*Department of Psychology, Dalhousie University, Halifax, Nova Scotia, Canada*
[r]*GlaxoSmithKline, Research Triangle Park, NC, USA*
[s]*Food and Drug Administration, Rockville, MD, USA*
[t]*Department of Physical Medicine and Rehabilitation, University of Washington, Seattle, WA, USA*
[u]*Elan Pharmaceuticals, San Diego, CA, USA*
[v]*Abbott Laboratories, Lake Forrest, IL, USA*
[w]*University of California San Diego, La Jolla, CA, USA*
[x]*US Department of Veterans Affairs, Washington, DC, USA*

Received 14 May 2003; received in revised form 1 August 2003; accepted 4 August 2003

Abstract

Objective. To provide recommendations for the core outcome domains that should be *considered* by investigators conducting clinical trials of the efficacy and effectiveness of treatments for chronic pain. Development of a core set of outcome domains would facilitate comparison and pooling of data, encourage more complete reporting of outcomes, simplify the preparation and review of research proposals and manuscripts, and allow clinicians to make informed decisions regarding the risks and benefits of treatment.

Methods. Under the auspices of the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT), 27 specialists from academia, governmental agencies, and the pharmaceutical industry participated in a consensus meeting and identified core outcome domains that should be *considered* in clinical trials of treatments for chronic pain.

Conclusions. There was a consensus that chronic pain clinical trials should assess outcomes representing six core domains: (1) pain, (2) physical functioning, (3) emotional functioning, (4) participant ratings of improvement and satisfaction with treatment, (5) symptoms and

* Corresponding author. Tel.: +1-206-616-2626; fax: +1-206-543-2958.
E-mail address: turkdc@u.washington.edu (D.C. Turk).

doi:10.1016/j.pain.2003.08.001

adverse events. (6) participant disposition (e.g. adherence to the treatment regimen and reasons for premature withdrawal from the trial). Although consideration should be given to the assessment of each of these domains, there may be exceptions to the general recommendation to include all of these domains in chronic pain trials. When this occurs, the rationale for not including domains should be provided. It is not the intention of these recommendations that assessment of the core domains should be considered a *requirement* for approval of product applications by regulatory agencies or that a treatment must demonstrate statistically significant effects for all of the relevant core domains to establish evidence of its efficacy.

Keywords: Chronic pain; Clinical trials; Assessment; Outcomes; Quality of life; Physical functioning; Emotional functioning; Adverse events; Global ratings

1. Introduction

Variability among clinical trials in outcome assessments has impeded evaluations of the efficacy and effectiveness of treatments for chronic pain, and the use of different outcome domains precludes meaningful comparisons among studies. One way to facilitate such evaluations would be through the use of a standard set of outcome domains. Although investigators may wish to augment a core set of domains with others that are specific to the situation or treatment being studied, use of a core set of outcome variables among studies would permit comparisons among different samples, treatments, and settings.

Development of a core set of outcome domains and measurement procedures would facilitate comparison and pooling of data while leaving investigators free to augment the core set with others of their choice. In addition, a core set of domains would encourage more complete investigation and reporting of relevant outcomes, so that investigators do not simply present a single outcome while ignoring others. Another advantage is that it would encourage development of cooperative multicenter projects, in which different centers agree to assess the core domains, in addition to any measures selected to evaluate specific research questions. A standard set of outcome domains would simplify the process of designing and reviewing research proposals, manuscripts, and published articles. Finally, published results of clinical trials with common outcome domains will allow clinicians to make more informed clinical decisions for each patient regarding the optimal treatment, especially with respect to its risks and benefits. Once core outcome domains for clinical trials are identified, the next step would be to select measures that meet appropriate psychometric standards (i.e. reliability, validity, responsiveness, appropriate, normative data).

To address the identification of core outcome domains, the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT, additional information concerning IMMPACT and its meetings can be found at immpact.org) convened a meeting to develop consensus recommendations for chronic pain clinical trials. There was agreement that the identification of specific measures would occur at a subsequent meeting. Other initiatives provide precedents for this undertaking, including Outcome Measures in Rheumatoid Arthritis Clinical Trials (OMERACT; Bellamy et al., 1997) and World Health Organization/International League of Associations for Rheumatology (WHO/ILARS; Brooks and Hochberg, 2001) in rheumatology, European Organization for Research and Treatment of Cancer (EORTC, Aaronson et al., 1993) and the Research Network of the European Association of Palliative Care (Caraceni et al., 2002) in oncology, and an international consortium of back pain researchers (Deyo et al., 1998). Although these other disease-specific initiatives were used to inform the discussion, the objective of the IMMPACT meeting was to develop a consensus on outcome domains that would transcend specific chronic pain syndromes. Our goal in this paper is to present the consensus recommendations from the first IMMPACT meeting for a core set of outcome domains that should be considered for all clinical trials of treatments for chronic pain.

2. Methods

2.1. Sponsorship

Abbott Laboratories, AstraZeneca, Elan Pharmaceuticals, Endo Pharmaceuticals Inc., GlaxoSmithKline, Novartis Pharmaceuticals, Ortho-McNeil Pharmaceutical Inc., Pfizer, and Purdue Pharma provided unrestricted educational grants to the University of Rochester Office of Professional Education to support a meeting and manuscript preparation.

2.2. Procedure

A meeting consisting of 27 people representing academia, governmental agencies, and the pharmaceutical industry was held on November 1–2, 2002. The participants attending the meeting were selected to represent health care disciplines that cover chronic pain broadly defined and included anesthesiology, biostatistics, clinical pharmacology, epidemiology, geriatrics, internal medicine, neurology, nursing, oncology, pediatric pain, physical medicine and rehabilitation, psychology, and rheumatology; all have research, clinical, or administrative expertise relevant to evaluating chronic pain treatment outcomes. In addition, representatives from the pharmaceutical industry who are engaged in chronic pain clinical trials and an attorney were included to provide specific expertise.

The process of the consensus meeting was semi-structured, with the first two authors leading discussions. Prior to the meeting, all participants were provided copies of a recent edited volume on pain assessment (Turk and Melzack, 2001), as well as four published clinical trials that are representative of chronic pain trials. Outcomes included in these studies were used to illustrate the diversity of domains examined in recent trials. The list of various domains generated by the participants was discussed and consensus was reached based on the results of the discussion and a formal vote.

The first two authors facilitated the consensus meeting and prepared the first draft of this paper. They were responsible for revising the manuscript and integrating the comments of the other authors. All authors reviewed the final draft and endorsed its publication.

3. General issues

To demonstrate the benefits of treatment, investigators must decide the appropriate endpoints for establishing both the statistical significance and the clinical importance of the effects of treatment. In a clinical trial of a treatment for chronic pain, pain reduction and safety are necessary outcome variables but they may not be sufficient for a comprehensive evaluation of the overall benefit or harm of treatment (Dionne and Witter, 2003). The complexity of chronic pain and its negative impact on diverse aspects of function is well established (e.g. Melzack and Wall, 1982). Thus, evaluation of the effectiveness of any treatment for chronic pain requires consideration of the assessment of multiple outcome domains to adequately characterize the impact of the intervention. Adverse events resulting from the treatment might outweigh the benefits of pain reduction, and pain reduction alone does not guarantee that physical or emotional functioning will improve.

The domains of importance in a clinical trial should match the purpose of the study, measure positive and negative outcomes of treatment, and be appropriate for the chronic pain syndrome studied and the specific characteristics of the sample (e.g. geriatric participants). Central issues involve the identification of outcome domains that are clinically meaningful and for which there are measures that are responsive and provide a comprehensive yet efficient evaluation of treatment response (Bellamy et al., 1997; Revicki and Ehreth, 1997).

4. Core outcome domains for chronic pain clinical trials

The authors recommend that each of the six core outcome domains listed in Table 1 should be *considered* in the design of all clinical trials of the efficacy and effectiveness of treatments for chronic pain. If one or more of these domains is not included in such a chronic pain clinical trial, the reasons for the exclusion should be justified a priori. Importantly, it is not the intention of these recommendations that assessment of these core domains should be considered a requirement for the approval of product applications by regulatory agencies or that a treatment must demonstrate statistically significant effects for all of the core domains to establish evidence of its efficacy. Rather, these recommendations are presented in an effort to promote collection and publication of standardized outcomes, which will allow for improved evidence-based comparisons and meta-analyses of chronic pain treatments. As noted above, there will be clinical trials in which these core assessment domains will require modification, for example, clinical trials in individuals with mild pain (in whom the impact of treatment on physical function and emotional distress may be less relevant than it is in patients with moderate or severe pain), single-dose studies in participants with a chronic pain syndrome, and clinical trials in the cognitively impaired and in infants and children.

Table 1
Core domains for clinical trials of chronic pain treatment efficacy and effectiveness

Pain
Physical functioning
Emotional functioning
Participant ratings of global improvement
Symptoms and adverse events
Participant disposition (including adherence to the treatment regimen and reasons for premature withdrawal from the trial)

Our recommendations are most applicable to clinical trials of treatments for chronic pain designed to evaluate efficacy or effectiveness, for example, what are termed Phase III and IV trials within the regulatory context (United States Department of Health and Human Services, 1997). These recommendations are made with the assumption that clinical trials will be conducted according to the principles of good clinical practice presented in the E6 Good Clinical Practice Consolidated Guidance of the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (United States Department of Health and Human Services, 1996).

4.1. Pain

There are several dimensions of pain that can be assessed in a clinical trial (e.g. intensity, location, specific descriptors and qualities). Most chronic pain clinical trials will also assess pain history, but these variables are more likely to be considered baseline characteristics or covariates.

It has often been assumed that chronic pain is highly associated with alterations in emotional and physical functioning and that reduction in pain will inevitably lead to improvement in function and satisfaction with treatment. This is not necessarily the case, and in many studies, pain and functioning are only modestly related. Moreover,

changes in pain severity may have only a variable relationship with participants' ratings of improvement and satisfaction (Dougados et al., 2002; Farrar et al., 2001; Dawson et al., 2002). Such data indicate that even though pain is typically considered the primary outcome in evaluating pain treatments, it is important to consider other outcomes in clinical trials.

4.2. Physical functioning

In addition to relieving clinical symptoms and prolonging survival, the objectives of health care intervention include improvement of functioning (Revicki et al., 2000). Thus, there is a need to assess multiple domains of functioning, including behavior, mood, and satisfaction (Ware, 1984; Revicki, 1993). *Quality of life* (QOL) is a term that refers to how a person feels and how he or she functions in daily life. Concerns with the all-encompassing nature of QOL in the evaluation of treatment outcomes have led a number of investigators to use a more circumscribed construct, *health-related quality of life* (HRQOL). HRQOL refers to those domains that are specifically related to health and that can be potentially influenced by the healthcare system (Varni et al., 1999; Seid et al., 2000). HRQOL outcomes are especially important for evaluating the impact of treatment on chronic diseases for which cure is not possible and therapy may be prolonged. Moreover, especially when treatment extends over long periods, it is critical to examine whether the benefits of symptom reduction are compromised by reductions in QOL resulting from adverse effects of treatment.

Several authors have argued that the assessment of QOL and HRQOL is problematic because of the lack of clear definitions and shared theoretical frameworks, which makes it difficult to determine whether a given scale is a valid measure (Faden and LePlege, 1992; Cella and Bonomi, 1995). The consensus of the authors is that two central components of existing HRQOL instruments, physical functioning and emotional functioning, are core domains that should be considered in all clinical trials of chronic pain treatments. This recommendation is supported by the results of studies in which exploratory and confirmatory factor analyses were used to identify the variables needed to comprehensively assess chronic pain participants, which suggested that three relatively independent domains—pain severity, physical functioning, and emotional functioning—are required to capture the multidimensionality of the pain experience (Mikail et al., 1993; De Gagné et al., 1995; Holroyd et al., 1999).

Measures of physical functioning evaluate diverse aspects of a participant's life, including the ability to carry out such daily activities as household chores, walking, work, travel, and self-care, as well as strength and endurance. A major decision to be made in assessing the impact of a treatment on physical functioning involves whether a generic or a disease-specific measure will be used (Stucki et al., 1995; Garratt et al., 2001). Disease-specific measures are designed to evaluate the impact of a specific condition (e.g. ability to wear clothing in participants with postherpetic neuralgia). Such specific effects of a disorder may not be assessed by a generic measure, and disease-specific measures may therefore be more likely to reveal clinically important improvement or deterioration in function that is a consequence of treatment. In addition, responses on disease-specific measures will generally not reflect the effects of co-morbid conditions on physical functioning, which may confound the interpretation of change occurring over the course of a trial when generic measures are used. Generic measures, however, make it possible to compare the physical functioning associated with a given disorder and its treatment with those of different conditions (Dworkin et al., 2001). Thus, the use of disease-specific and generic measures in combination facilitates the achievement of both sets of objectives (Patrick and Deyo, 1989).

Different levels of analysis can be used to conceptualize the core outcome domain of physical functioning. For example, activities of daily living such as performing self-care behaviors (e.g. bathing and dressing) can be distinguished from social-role functioning. The consensus of the meeting was that these two levels of activities should be differentiated with activities of daily living being more fundamental than engaging in social activities. Consequently, there was agreement that the effect of the treatment on the ability of the participant to perform specific physical tasks or the reduction in the interference of the pain in the participant's ability to engage in routine, daily physical activities should be treated as a core domain, whereas the impact of treatment on alteration in social functioning was considered a supplemental domain.

4.3. Emotional functioning

The results of numerous studies suggest that chronic pain is often associated with emotional distress, particularly depression, anxiety, anger, and irritability (e.g. Fernandez and Turk, 1995; Banks and Kerns, 1996; Robinson and Riley, 1999). Emotional functioning as reflected in emotional distress, is not intended to be synonymous with a psychiatric diagnosis or disorder, but is rather meant to refer to distressed mood more generally. The consensus of the participants was that the assessment of emotional functioning should be considered a core outcome in chronic pain clinical trials. Although it is difficult to interpret changes in emotional functioning because of the many factors that contribute, this domain is central in people's assessments of their well-being and satisfaction with life and the authors recommend that it should be considered a core outcome domain in clinical trials of treatments for chronic pain.

4.4. Participant ratings of global improvement and satisfaction with treatment

Assessments of individual outcome domains such as pain and physical and emotional functioning may not adequately characterize the participant's expectations about the treatment, overall assessment of treatment, and the meaningfulness to the participant of any improvement (or worsening). Global evaluations by participants in clinical trials of the benefits of treatment reflect not only the magnitude of the changes in these outcomes and feelings about treatment delivery, but also the personal importance that these outcomes have for participants. Such perceptions of the importance of treatment-associated changes often differ considerably from those of health care professionals (Lipton and Stewart, 1999), and the value and significance of therapeutic changes differ greatly among participants and are important determinants of their treatment satisfaction.

The use of participants' overall evaluation of treatment in clinical trials is controversial. A substantial amount of confusion about this group of outcomes is generated by vastly different meaning applied to terms such as 'patient satisfaction' and 'impression of change'. In addition, many such assessments are based on rating a single item, and it is not possible to establish the internal consistency of one rating. In addition, global impressions of improvement may fail to detect important changes (e.g. Just et al., 1999). Furthermore, the judgment of change requires participants to assess both their present and initial state and then perform what may be an unreliable mental subtraction; because participants may be unable to recall their initial state, their ratings may be based on an 'implicit theory' of change beginning with their present state and working backward (Ross, 1989). However, if a treatment is associated with severe adverse effects, the participant may not need to remember baseline pain to rate satisfaction with treatment. In addition to problems of memory recall, participants' global impressions may be influenced by systematic biases such as the desire to please health care providers (e.g. demand characteristics). Participants' efforts to comply with their perceptions of provider expectations might also contribute to global judgments beyond the actual balancing of perceived benefits against accompanying negative effects. Despite the necessity for care in the use of participant global assessments, the results of recent research provide support for their validity (e.g. Fischer et al., 1999; Collins et al., 2001; Farrar et al., 2001).

Ultimately, participants decide whether the positive attributes of a treatment outweigh its negative aspects, and this is an important determinant of whether they adhere to and continue with treatment. Willingness to continue with the treatment regimen may be viewed as a gross indication of participant satisfaction. A more systematic approach is to ask participants to rate their degree of satisfaction. Such ratings permit a range of satisfaction beyond the dichotomous behavior of withdrawal from a protocol. Participant ratings of improvement and satisfaction with treatment provide unique information in outcomes assessment in clinical trials because they may allow an integration of the benefits of treatment and adverse events and other costs from within the participant's personal perspective. The authors therefore recommend that at least one rating of global improvement should be considered for inclusion in all chronic pain clinical trials, but must be carefully constructed to capture the relevant data.

4.5. Symptoms and adverse events

Many participants will experience symptoms and adverse events associated with their illness and pharmacologic treatment. The importance of monitoring adverse events has long been recognized as an essential component of all therapeutic clinical trials (Anderson and Testa, 1994). Therapies, such as the drugs that relieve pain, have a variety of effects, and these cannot only cause discomfort but also may potentially impair physical and emotional function and exacerbate co-morbid symptoms, which thereby may potentially offset the therapeutic benefit (Croog et al., 1986). Max and Laska (1991) have noted that common analgesic adverse events (e.g. gastrointestinal distress, sedation, depression) can limit the dosage that can be realistically prescribed. Moreover, side effect burden plays an important role in treatment adherence (Anderson et al., 1999). Participants may view adverse events as sufficiently noxious to discontinue treatment or limit dosage, and the overall benefit of treatment may therefore be reduced. A major challenge in developing analgesic drugs is determining an optimal dosage (i.e. one that minimizes adverse events and maximizes pain relief and functional improvement).

The onset of new diseases and initiation of new treatments during a clinical trial complicates assessments of symptoms and adverse events. When initiated during a trial, concomitant treatments (e.g. drugs, physical therapy, psychological therapy, nerve blocks) are often protocol violations. Participant disease is a baseline characteristic or covariate when present at the beginning of a trial but is an adverse event when it emerges or worsens. The risk of addiction has attracted considerable attention in the evaluation of analgesic drugs. Addiction is a neurobiologic disease, and if it occurs during a trial it should be considered an adverse event but when it is present at the beginning of a trial it is a baseline characteristic. As a caution, we note that addiction is not the same as physical dependence or tolerance. Physical dependence is a pharmacologic consequence of a drug characterized by the occurrence of a withdrawal syndrome following abrupt discontinuation of the substance or the administration of an antagonist. Tolerance refers to a physiologic state in which increased dosages of a substance are required to sustained a desired effect.

Assessment of the percentages of participants experiencing adverse events based on passive capture is standard in clinical trials; however, assessments of their severity and

importance to participants are much less common, although this may provide valuable information (Katz, 2002). The authors recommend that the prospective assessment of symptoms present at the onset of a trial and symptoms and adverse events that emerge during the trial is a core outcome domain that should be included in all chronic pain clinical trials, and that the strategy used to assess these events should include participant ratings of their presence, severity, change, and importance.

4.6. Participant disposition

Following the recommendations of the CONSORT statement (Consolidated Standards of Reporting Trials guidelines, CONSORT; Begg et al., 1996; Moher et al., 2001), all participants screened for a clinical trial should be carefully described with respect to the proportion who are ultimately enrolled, and why those who were not enrolled were not. Detailed information should be provided regarding the extent and reasons for treatment non-adherence, prohibited concomitant medications and all other protocol deviations that may impact the interpretation of the trial results, treatment modification, premature participant withdrawal from the trial, and loss to follow-up. Investigators should report the number of withdrawals related to each of the symptoms and adverse events identified in each of the treatment groups. This detailed characterization of participant disposition is the sixth core domain that should be assessed in all clinical trials of chronic pain treatment.

To be effective, a treatment must have a beneficial effect on the symptom or disease being treated and the participant must adhere to the treatment regimen (Turk and Rudy, 1991). The most potent analgesic may demonstrate less than its potential benefit if participants in a clinical trial fail to use the medication in the manner prescribed, are unable to tolerate a fully effective dose, or drop out of the trial due to unacceptable adverse events or inadequate pain relief. Furthermore, the benefit of the treatment being studied may be obscured if participants receive any treatments that are not allowed in the protocol.

The dosage and duration of all treatments received by participants during the clinical trial must be recorded, not only the treatment being investigated, but also all concomitant treatments. Treatments initiated during the trial often reflect inadequate pain relief or the presence of distressing or uncontrolled adverse events (the use of rescue medications and changes in concomitant medication use may be justifiable as pain outcome measures when specified in the protocol). Assessments of the use of rescue and prohibited medications and alterations in prescribed treatment due to adverse events and symptoms must be considered in evaluating the results of chronic pain clinical trials.

To evaluate whether side effects or other factors have compromised the double-blind in a clinical trial, it is important to assess subjects' and investigators' guesses of which treatment was administered. The reasons for the specific guesses should also be assessed, because these can have different implications for interpretation of the results, for example, unblinding occurring because of the effectiveness of the active treatment or because of its side effects (Moscucci et al., 1987).

Table 2
Supplemental domains for chronic pain clinical trials

Role functioning (i.e. work and educational activities)
Interpersonal functioning (i.e. relationships and activities with family, friends, and others)
Pharmacoeconomic measures and health care utilization
Biological markers (e.g. assessments based on quantitative sensory testing, imaging, genetic markers, pharmacogenomics, and punch skin biopsy)
Coping
Clinician or surrogate ratings of global improvement
Neuropsychological assessments of cognitive and motor function
Suffering and other end-of-life issues

5. Supplemental outcome domains

There are many other outcome domains that can be considered in the design of pain clinical trials depending on the specific research question. Supplemental assessment domains may be included in a clinical trial without a hypothesis that they will change and without the trial having adequate power to test the hypothesis that they will respond to treatment. Table 2 contains a list of eight supplemental outcome domains that might be considered in the design of chronic pain clinical trials.

6. Conclusions

The core outcome domains specified in these IMMPACT consensus recommendations—pain, physical functioning, emotional functioning, participant ratings of global improvement and satisfaction with treatment, symptoms and adverse events, and participant disposition—are generally consistent with the recommendations for arthritis clinical trials from OMERACT-III (Bellamy et al., 1997) and WHO/ILARS (Brooks and Hochberg, 2001). Pain, physical function, participant global assessment, and imaging studies are the core outcome domains specified in the OMERACT guidelines, and the first three of these domains are included in the present recommendations.

The objective of the first IMMPACT consensus meeting was to establish recommendations for clinical trials of chronic pain treatment. Imaging studies were not considered because they have limited relevance to the assessment of outcome in many chronic pain syndromes. In addition to the three domains that overlap with the OMERACT guidelines, the authors consider emotional functioning a core outcome domain because of its well-established and clinically important relationships with chronic pain. Symptoms and

adverse events have been included as a core domain to emphasize that comprehensive assessment of the health burdens that often accompany treatment is necessary to achieve the key purpose of clinical trials—assessment of the risk-benefit balance. The recommendation that participant disposition is a core outcome domain is consistent with the CONSORT guidelines (Begg et al., 1996; Moher et al., 2001), although we have emphasized that reports of participant disposition should be accompanied by detailed explanations of withdrawals, non-adherence, and protocol violations.

A legitimate concern for any clinical trial is participant burden. Assessment of the six core domains will inevitably require more effort from participants than simply assessing pain reduction as the sole end-point of importance. However, it is important to emphasize that there are reasonably brief measures available that are capable of capturing the domains described above. Attention toward identifying measures that have demonstrated appropriate psychometric properties with the least participant burden will be the focus of the second IMMPACT consensus recommendations. Those who are designing clinical trials for chronic pain will need to balance the importance of assessing the core domains against the added participant burden.

The authors believe that investigators designing and conducting clinical trials of chronic pain treatment efficacy and effectiveness should consider each of the six core domains listed in Table 1 and discussed in this paper. It is important to emphasize, however, that we are not suggesting that positive results must be obtained for all of the core domains for the treatment to be deemed efficacious. Also we would like to emphasize the word considered. These core domains should be considered and are not mandatory because it is possible that there are specific trials for which one or more of these domains might not be relevant. In such instances, our recommendation is that investigators should acknowledge that they have considered each outcome domain and provide the rationale when they decide not to include assessment of a particular domain. Of course, there are many supplemental outcome domains that can be included in a chronic pain clinical trial (see Table 2), and we expect that the core outcome domains will be supplemented by assessment of additional domains that are required to evaluate a specific treatment (or that the investigator wishes to include for exploratory purposes).

Numerous outcome measures related to the recommended core domains have appeared in the research literature (e.g. Benzon et al., 1994; McDowell and Newell, 1996; Turk and Melzack, 2001). Selection of specific measures of each of the core and supplemental outcome domains from the many available should be based on reliability, validity, responsiveness to change, feasibility and practicality within the clinical trial setting (e.g. participant and investigator burden, need for special training), availability of normative data and linguistically and culturally validated versions, mode of administration, and appropriateness to study objectives and the participant population and treatment being investigated (Dworkin et al., 2001). Future IMMPACT recommendations will focus on identifying specific measures within the six core outcome domains that have the most favorable characteristics and the widest range of applicability, methods for determining the clinical importance to patients of changes in these measures, and strategies for selecting primary endpoints and combining multiple endpoints in assessments of treatment efficacy and effectiveness. The use of standard outcome assessments has the potential to greatly enhance the validity, comparability, and clinical applicability of clinical trials of chronic pain treatments.

Academic, health care, and pharmaceutical industry investigators who conduct clinical trials, the government and private organizations that provide funding for many such studies, and the government regulatory agencies that review this research and ultimately approve new therapies for the public all share a commitment to identifying treatments for chronic pain that are more effective and have fewer adverse effects than those currently available. These different groups, however, sometimes have different goals, contrasting ideologies, and separate constituencies with particular interests in clinical trials. Although unsystematic efforts to bring these different individuals together have occurred in various medical specialties, much more can and should be done to enhance mutual understanding and promote creativity in the development and investigation of improved treatment approaches (Klein et al., 2002). The authors hope that IMMPACT and the recommendations made in this paper will provide an example of the value of such collaborative efforts among academia, government, and industry. The ultimate goal of such efforts should be to advance the science of chronic pain clinical trials and thereby provide improved treatments for patients suffering from chronic pain.

7. Disclaimer

The views expressed in this paper are those of the authors. No official endorsement by the US Department of Veterans Affairs, US Food and Drug Administration, US National Institutes of Health, or any of the pharmaceutical companies that provided unrestricted educational grants to the University of Rochester Office of Professional Education should be inferred.

References

Aaronson NK, Ahmedzai S, Bergman B. The European Organization for Research and Treatment of Cancer QLQ-c30: a quality-of-life

instrument of use in international clinical trials in oncology. J Natl Cancer Inst 1993;85:365–76.
Anderson RB, Hollenberg NK, Williams GH. Physical symptoms distress index: a sensitive tool to evaluate the impact of pharmacological agents. Arch Int Med 1999;159:693–700.
Anderson RB, Testa MA. Symptom distress checklists as a component of quality of life measurement: comparing prompted reports by patient and physician with concurrent adverse event reports via the physician. Drug Inf J 1994;28:89–114.
Banks SM, Kerns RD. Explaining high rates of depression in chronic pain: a diathesis-stress framework. Psychol Bull 1996;119:95–110.
Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz KF, Simel D, Stroup DF. Improving the quality of reporting of randomized controlled trials: the CONSORT statement. J Am Med Assoc 1996;276:637–9.
Bellamy N, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P, Altman R, Brandt K, Dougados M, LeQuesne M. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis: consensus development at OMERACT III. J Rheumatol 1997;24:799–802.
Benzon RA, Donnelly MA, Simpson RL, Simeon GP, Tilson HH. Quality of life bibliography and indexes. Qual Life Res 1994;6:547–68.
Brooks P, Hochberg M, for ILAR and OMERACT. Outcome measures and classification criteria for the rheumatic diseases: a compilation of data from OMERACT (Outcome Measures for Arthritis Clinical Trials), ILAR (International League of Associations for Rheumatology), regional leagues and other groups. Rheumatology 2001;40: 896–906.
Caraceni A, Cherny N, Fainsinger R, Kaasa S, Poulain P, Radbruch L, De Conno F, Steering Committee of EAPC Research Network. Pain measurement tools and methods in clinical research in palliative care: recommendations of an expert working group of the European Association of Palliative Care. J Pain Symptom Mgmt 2002;23: 239–55.
Cella D, Bonomi AE. Measuring quality of life: update. Oncology 1995; 1995:47–60.
Collins SL, Edwards J, Moore RA, Smith LA, McQuay HJ. Seeking a simple measure of analgesia for mega-trials: is a single global assessment good enough? Pain 2001;91:189–94.
Croog SH, Levine S, Testa MA, Brown B, Bulpitt CJ, Jenkins CD. The effects of antihypertensive therapy on quality of life. N Engl J Med 1986;314:1657–64.
Dawson R, Spross JA, Jablonski ES, Hoyer DR, Sellers DE, Solomon MZ. Probing the paradox of patients' satisfaction with inadequate pain management. J Pain Symptom Mgmt 2002;23:211–20.
De Gagné TA, Mikail SF, D'Eon JL. Confirmatory factor analysis of a 4-factor model of chronic pain evaluation. Pain 1995;60:195–202.
Deyo RA, Battie M, Beurskens AHHM, Bombardier C, Croft P, Koes B, Malmivaara A, Roland M, Von Korff M, Waddell G. Outcome measures for low back pain research: a proposal for standardized use. Spine 1998;23:2003–13.
Dionne R, Witter J. NIH-FDA analgesic drug development workshop: translating scientific advances into improved pain relief. Clin J Pain 2003;19:139–47.
Dougados M, LeClaire P, van der Heijde D, Bloch DA, Bellamy N, Altman RD. A report of the Osteoarthritis Research Society International Standing Committee for Clinical Trials Response Criteria Initiative. Osteoarthritis Cartilage 2002;8:395–403.
Dworkin RH, Nagasako EM, Heuzel RD, Farrar JT. Handbook of pain assessment. In: Turk DC, Melzack R, editors. Assessment of pain and pain-related quality of life in clinical trials, 2nd ed. New York: Guilford Press; 2001. p. 659–92.
Faden R, LePlege A. Assessing quality of life: moral implications for clinical practice. Med Care 1992;30(Suppl. 5):MS166–75.
Farrar JT, Young Jr JP, LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain 2001;94:149–58.
Fernandez E, Turk DC. The significance of anger in the experience of chronic pain. Pain 1995;61:165–75.
Fischer D, Stewart AL, Bloch DA, Lorig K, Laurent D, Holman H. Capturing the patient's view of change as a clinical outcome measure. J Am Med Assoc 1999;282:1157–62.
Garratt AM, Moffett JK, Farrin AJ. Responsiveness of generic and specific measures of health outcome in low back pain. Spine 2001;26:71–7.
Holroyd KA, Malinoski P, Davis MK, Lipchik GL. The three dimensions of headache impact: pain, disability and affective distress. Pain 1999;83: 571–8.
Just N, Ciccone DS, Bandilla EB, Wu W-H. Global impressions versus validated measures of treatment effectiveness in patients with chronic nonmalignant pain. Rehab Psychol 1999;44:194–207.
Katz N. The impact of pain management on quality of life. J Pain Symptom Mgmt 2002;24:S38–47.
Klein DF, Thase ME, Endicott J, Adler L, Glick I, Kalali A, Leventer S, Mattes J, Ross P, Bystritsky A. Improving clinical trials American Society of Clinical Psychopharmacology recommendations. Arch Gen Psychiatry 2002;59:272–8.
Lipton R, Stewart W. Acute migraine therapy: do doctors understand what patients with migraine want from therapy? Headache 1999;39(Suppl. 2):S20–6.
Max MB, Laska EM. Single-dose analgesic comparisons. In: Max M, Portenoy R, Laska E, editors. Advances in pain research and therapy, vol. 18, New York: Raven Press; 1991. p. 55–95.
McDowell I, Newell C. Measuring health: a guide to rating scales and questionnaires, 2nd ed. New York: Oxford University Press; 1996.
Melzack R, Wall PD. The challenge of pain. New York: Basic Books; 1982.
Mikail SF, DuBreuil S, D'Eon JL. A comparative analysis of measures used in the assessment of chronic pain patients. Psychol Assess 1993;5:117–20.
Moher D, Schulz K, Altman DG, for the CONSORT Group. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. Ann Int Med 2001;134:657–62.
Moscucci M, Byrne L, Weintraub M, Cox C. Blinding, unblinding, and the placebo effect: an analysis of patients' guesses of treatment assignment in a double-blind clinical trial. Clin Pharmacol Ther 1987;41:259–65.
Patrick DL, Deyo RA. Generic and disease-specific measures in assessing health status and quality of life. Med Care 1989;27:S217–32.
Revicki DA. Health care technology assessment and health-related quality of life. In: Banta D, Luce BR, editors. Health care technology and its assessment: an international perspective. New York: Oxford University Press; 1993. p. 114–31.
Revicki DA, Ehreth JL. Health-related quality of life assessment and planning for the pharmaceutical industry. Clin Ther 1997;19:1101–15.
Revicki DA, Osoba D, Fairclough D, Barofsky I, Berzon R, Leidy NK, Rothman M. Recommendations on health-related quality of life research to support labeling and promotional claims in the United States. Qual Life Res 2000;8:887–900.
Robinson ME, Riley JL. The role of emotion in pain. In: Gatchel RJ, Turk DC, editors. Psychosocial factors in pain: clinical perspectives. New York: Guilford Press; 1999. p. 74–88.
Ross M. Relation of implicit theories to the construction of personal histories. Psychol Rev 1989;96:341–7.
Seid M, Varni J, Jacobs J. Pediatric health-related quality-of-life measurement technology: intersections between science, managed care, and clinical care. J Clin Psychol Med Settings 2000;7:17–27.
Stucki G, Liang MH, Fossel AH, Katz JN. Relative responsiveness of condition-specific and generic health status measures in degenerative lumbar stenosis. J Clin Epidemiol 1995;48:1369–78.
Turk DC, Melzack R, editors. Handbook of pain assessment, 2nd ed. New York: Guilford Press; 2001.
Turk DC, Rudy TE. Neglected factors in chronic pain treatment outcome studies—relapse, noncompliance, and adherence enhancement. Pain 1991;44:528.
United States Department of Health and Human Services. Guidance for industry: E6 good clinical practice: consolidated guidance. Washington,

DC: Center for Drug Evaluation and Research, Food and Drug Administration; 1996 (www.fda.gov/cder/guidance/959fnl.pdf).

United States Department of Health and Human Services. Food and Drug Administration (Docket No. 97D-0188): International Conference on Harmonisation; guidance on general considerations for clinical trials. Federal Register 1997;62:66113–9 (www.fda.gov/cder/guidance/1857fnl.pdf).

Varni JW, Seid M, Kurtin PS. Pediatric health-related quality of life measurement technology: a guide for healthcare decision makers. JCOM 1999;6:33–40.

Ware JE. Methodological consideration in the selection of health status assessment procedures. In: Wenger N, Mattson M, Furberg C, Elinson J, editors. Assessment of quality of life in clinical trials of cardiovascular therapies. Greenwich, CT: La Jacq; 1984. p. 87–111.

PAIN

Pain 113 (2005) 9–19

www.elsevier.com/locate/pain

Topical Review and Recommendations

Core outcome measures for chronic pain clinical trials: IMMPACT recommendations

Robert H. Dworkin[a,*], Dennis C. Turk[b], John T. Farrar[c], Jennifer A. Haythornthwaite[d], Mark P. Jensen[b], Nathaniel P. Katz[e], Robert D. Kerns[f], Gerold Stucki[g], Robert R. Allen[h], Nicholas Bellamy[i], Daniel B. Carr[j], Julie Chandler[k], Penney Cowan[l], Raymond Dionne[m], Bradley S. Galer[n], Sharon Hertz[o], Alejandro R. Jadad[p], Lynn D. Kramer[q], Donald C. Manning[r], Susan Martin[s], Cynthia G. McCormick[t], Michael P. McDermott[u], Patrick McGrath[v], Steve Quessy[w], Bob A. Rappaport[o], Wendye Robbins[x], James P. Robinson[b], Margaret Rothman[y], Mike A. Royal[z], Lee Simon[o], Joseph W. Stauffer[aa], Wendy Stein[ab], Jane Tollett[ac], Joachim Wernicke[ad], James Witter[o]

[a]*Department of Anesthesiology, University of Rochester School of Medicine and Dentistry, Rochester, NY 14642, USA*
[b]*University of Washington, Seattle, WA, USA*
[c]*University of Pennsylvania, Philadelphia, PA, USA*
[d]*Johns Hopkins University, Baltimore, MD, USA*
[e]*Harvard University, Boston, MA, USA*
[f]*Yale University, New Haven, CT, USA*
[g]*University of Munich, Munich, Germany*
[h]*AstraZeneca, Wilmington, DE, USA*
[i]*University of Queensland, Brisbane, Qld, Australia*
[j]*Tufts University, Boston, MA, USA*
[k]*Merck & Co., Blue Bell, PA, USA*
[l]*American Chronic Pain Association, Rocklin, CA, USA*
[m]*National Institute of Dental and Craniofacial Research, Bethesda, MD, USA*
[n]*Endo Pharmaceuticals Inc., Chadds Ford, PA, USA*
[o]*Food and Drug Administration, Rockville, MD, USA*
[p]*University Health Network and University of Toronto, Toronto, Ont., Canada*
[q]*Purdue Pharma, Stamford, CT, USA*
[r]*Novartis Pharmaceuticals Corporation, East Hanover, NJ, USA*
[s]*Pfizer Global Research and Development, Ann Arbor, MI, USA*
[t]*National Institute of Neurological Disorders and Stroke, Bethesda, MD, USA*
[u]*University of Rochester, Rochester, NY, USA*
[v]*Dalhousie University, Halifax, NS, Canada*
[w]*GlaxoSmithKline, Research Triangle Park, NC, USA*
[x]*NeurogesX, San Carlos, CA, USA*
[y]*Johnson & Johnson, Raritan, NJ, USA*
[z]*Elan Pharmaceuticals, San Diego, CA, USA*
[aa]*Abbott Laboratories, Lake Forest, IL, USA*
[ab]*University of California San Diego, San Diego, CA, USA*
[ac]*US Department of Veterans Affairs, Washington, DC, USA*
[ad]*Eli Lilly & Co., Indianapolis, IN, USA*

Received 10 March 2004; received in revised form 1 September 2004; accepted 13 September 2004

* Corresponding author. Tel.: +1 585 275 3524; fax: +1 585 473 5007.
E-mail address: robert_dworkin@urmc.rochester.edu (R.H. Dworkin).

doi:10.1016/j.pain.2004.09.012

Keywords: Chronic pain; Clinical trials; Randomized controlled trials; Assessment; Outcomes; Health-related quality of life; Physical functioning; Emotional functioning; Global ratings; Adverse events

1. Introduction

Many patients with chronic pain do not obtain adequate relief or experience unacceptable side effects from existing treatments. Moreover, even when clinical trials report positive outcomes, the long-term benefits of these treatments have not been demonstrated. Efforts to develop treatments that provide improved outcomes are therefore a priority for pain research. Because variability in outcome measures across clinical trials hinders evaluations of the efficacy and effectiveness of treatments, the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) has recently recommended that 6 core outcome domains should be *considered* when designing chronic pain clinical trials. These 6 core outcome domains were: (1) pain; (2) physical functioning; (3) emotional functioning; (4) participant ratings of improvement and satisfaction with treatment; (5) symptoms and adverse events; and (6) participant disposition (Turk et al., 2003).

The benefits of adopting these core outcome domains in clinical research on chronic pain would be augmented by the identification of optimal measures for assessing them. Such core outcome measures could be supplemented by measures specific to the situation or treatment being studied. Use of a standard set of outcome measures for chronic pain clinical trials would facilitate the process of developing research protocols, encourage development of multi-center projects in which all participating facilities agree to include these measures, provide a basis for determining the treatment outcomes that constitute clinically important differences, permit pooling of data from different studies, and provide a basis for meaningful comparisons among treatments of the clinical importance of their outcomes, particularly through systematic reviews (Jadad, 1998; Jadad and Cepeda, 2000).

IMMPACT-II was convened to develop consensus recommendations for specific measures of each of the IMMPACT core outcome domains. Although there have been recent attempts to recommend outcome measures for specific chronic pain conditions—including osteoarthritis (Bellamy et al., 1997), low back pain (Deyo et al., 1998), and neuropathic pain (Cruccu et al., 2004)—the only previous attempt to identify specific treatment outcome measures applicable to diverse chronic pain conditions was published over fifteen years ago (Williams, 1988). The objective of the present article is to present consensus recommendations for specific measures of each of the IMMPACT core outcome domains.

2. Consensus meeting procedure

The IMMPACT-II meeting was held on April 11–12, 2003 and included 35 participants from academia, governmental agencies, a self-help organization, and the pharmaceutical industry. The participants were selected on the basis of their research, clinical, or administrative expertise relevant to the design and evaluation of chronic pain treatment outcomes. Literature reviews of measures of the IMMPACT core outcome domains were commissioned specifically for the IMMPACT-II meeting and distributed to participants prior to the meeting. These reviews focused on measures that could be used in trials of all chronic pain conditions and did not examine measures that were specific to certain types of chronic pain. These background literature reviews and the slide presentations delivered at the meeting are available on the IMMPACT-II page at www.immpact.org/meetings.html. They should be consulted for detailed reviews and discussions of the measures that were considered, the evidence on which the present recommendations are based, and the reasons for selection or rejection of specific measures.

Among the criteria used in evaluating potential core outcome measures were: (1) appropriateness of the measure's content and conceptual model; (2) reliability; (3) validity; (4) responsiveness; (5) interpretability; (6) precision of scores; (7) respondent and administrator acceptability; (8) respondent and administrator burden and feasibility; (9) availability and equivalence of alternate forms and methods of administration (e.g. self-report, interviewer); and (10) availability and equivalence of versions for different cultures and languages (Fitzpatrick et al., 1998; Scientific Advisory Committee of the Medical Outcomes Trust, 2002). Responsiveness has been defined and assessed in numerous ways, but it most often refers to the ability of a measure to detect changes over time (Guyatt et al., 1987; Terwee et al., 2003). With respect to clinical trials, responsiveness has also referred to the ability of a measure to distinguish between treatments, in particular, between an active/experimental treatment and a placebo/control treatment. Although Hays and Hadorn (1992) have noted that responsiveness is a component of validity, the authors considered responsiveness a separate attribute of outcome measures because of its pivotal role in clinical trials.

In evaluating the extent to which the various measures reviewed in the background presentations fulfilled these criteria, appropriateness of content, reliability, validity, responsiveness, and participant burden were given the greatest weight. In particular, measures for which published information on these specific criteria were lacking were not recommended, and when such information was available for two or more relevant measures, recommendations were primarily based on comparisons of these five attributes. It is important to emphasize that even though basic information on reliability and validity is usually available for measures

that have been used in studies of patients with chronic pain, information on other important attributes of these measures is often lacking. The absence of data relevant to a measure's responsiveness, for example, must therefore be carefully distinguished from the availability of data that demonstrate its lack of responsiveness. Unfortunately, the absence of evidence is much more common than clear evidence of limitations for most of the criteria considered in evaluating outcome measures for chronic pain clinical trials.

Reliability, validity, and responsiveness can be condition or context specific and are not invariant properties of a measure. Although the authors considered evidence of the generalizability of these attributes to diverse chronic pain syndromes, in circumstances in which such data are lacking, it is important to evaluate the applicability of the measure to the chronic pain syndrome being investigated.

3. Core outcome measures for chronic pain clinical trials

The core outcome measures listed in Table 1 should be *considered* in the design of all clinical trials of the efficacy and effectiveness of treatments for any type of chronic pain. It is not the intention of these recommendations that use of these measures should be considered a requirement for approval of applications by regulatory agencies or that treatments must demonstrate statistically significant or clinically important benefits with all of these measures to establish evidence of efficacy or effectiveness. There may be circumstances in which use of some or all of these core outcome measures will not be appropriate, for example, in clinical trials in cognitively impaired individuals or in infants. As was true of the IMMPACT recommendations for core outcome domains (Turk et al., 2003), the present recommendations are most applicable to clinical trials to determine the efficacy or effectiveness of treatments for chronic pain and are made with the assumption that these trials will be conducted in accord with the principles of good clinical practice (International Conference on Harmonisation, 1996a; United States Department of Health and Human Services, 1997).

Table 1
Recommended core outcome measures for clinical trials of chronic pain treatment efficacy and effectiveness

Pain
11-point (0–10) numerical rating scale of pain intensity
Usage of rescue analgesics
Categorical rating of pain intensity (none, mild, moderate, severe) in circumstances in which numerical ratings may be problematic
Physical functioning (either one of two measures)
Multidimensional Pain Inventory Interference Scale
Brief Pain Inventory interference items
Emotional functioning (at least one of two measures)
Beck Depression Inventory
Profile of Mood States
Participant ratings of global improvement and satisfaction with treatment
Patient Global Impression of Change
Symptoms and adverse events
Passive capture of spontaneously reported adverse events and symptoms and use of open-ended prompts
Participant disposition
Detailed information regarding participant recruitment and progress through the trial, including all information specified in the CONSORT guidelines

3.1. Pain

There are various aspects of pain that can change as a result of treatment, and the results of reviews of the literature on pain assessment in adults (Jensen, 2003; Jensen and Karoly, 2001) support the recommendation that measures of pain intensity, the use of rescue treatments, pain quality, and the temporal components of pain should be considered when assessing pain outcomes. Self-report measures provide the 'gold standard' in assessing pain outcomes because they reflect the inherently subjective nature of pain, but they should be supplemented by careful assessments of the use of rescue treatments. Depending on the specific objectives of the clinical trial, other approaches to assessing pain can be considered, for example, overt expressions of pain and distress ('pain behaviors'; Keefe et al., 2001) and surrogate endpoints such as imaging measures.

3.1.1. Pain intensity

For most clinical trials of chronic pain treatments, a measure of pain intensity will provide the primary outcome measure. Each of the commonly used methods of rating pain intensity, including visual analogue scales (VAS), numerical rating scales (NRS), and verbal rating scales (VRS) are reliable and valid, and no one scale consistently demonstrates greater responsiveness in detecting improvements associated with pain treatment (Jensen and Karoly, 2001). However, there are important differences among VAS, NRS, and VRS measures of pain intensity with respect to lost data from patients failing to complete the measure correctly, patient preference, ease of data recording, and ability to administer the measure by telephone or with electronic diaries. VRS and NRS measures tend to be preferred over VAS measures by patients. Furthermore, VAS measures usually demonstrate greater amounts of missing and incomplete data than NRS measures, presumably because NRS measures are less abstract and easier to understand. Greater difficulty completing VAS measures is associated with increased age and opioid intake (Jensen and Karoly, 2001). Cognitive impairment has been shown to be associated with inability to complete NRS ratings of pain intensity (Jensen and Karoly, 2001). Patients who are unable to complete NRS ratings may be able to complete VRS pain ratings. There will, of course, be circumstances when self-reports of pain will be impossible and in these instances alternatives (e.g. observations of behavior, surrogate ratings) will have to be considered.

On the basis of a review of the literature on pain measures prepared for the IMMPACT-II consensus meeting (Jensen, 2003) and discussions among the participants, an 11-point (i.e. 0–10) NRS measure of pain intensity is recommended as a core outcome measure in clinical trials of chronic pain treatments. In order to facilitate consistency among studies, the authors recommend that the specific format of this rating should include presentation of the numbers from 0 to 10, with 0 meaning 'No pain' and '10' meaning 'Pain as bad as you can imagine,' accompanied by the instructions "Please rate your pain by indicating the number that best describes your pain on average in the last 24 h" (Cleeland and Ryan, 1994). Depending on the specific aims and design of the clinical trial, pain during the past week can also be assessed using this scale, as could pain 'at its worst' or pain 'at its least'.

Investigators should also routinely consider including a VRS measure of pain intensity (none, mild, moderate, severe) as an additional pain outcome measure. Doing so makes it possible to compare the results of a clinical trial with the many studies, especially of acute pain, that have used such VRS measures. In addition, use of a VRS measure of pain intensity should limit the amount of missing data that results from some study participants having difficulty completing the primary NRS measure.

There are clinical conditions for which reliable, valid, and responsive measures of pain intensity that do not use an NRS are routinely used (e.g. the Western Ontario and McMaster Universities Osteoarthritis Index [WOMAC] VAS ratings in studies of patients with osteoarthritis; Bellamy et al., 1988). These circumstances should be distinguished from those for which no such measures exist and NRS ratings of pain intensity are recommended. When other measures of pain intensity are used, it may be useful to also administer NRS ratings to compare with other diseases or treatments.

In addition to analyzing and reporting absolute changes in pain intensity, it is recommended that the percentages of patients obtaining reductions in pain intensity from baseline of at least 30% be reported when an NRS (or VAS) has been used in a chronic pain clinical trial. This recommendation is primarily based on the results of an analysis of the relationships between changes in pain intensity and patient reports of overall improvement in ten clinical trials of chronic pain in patients with diverse diagnoses (Farrar et al., 2001). Importantly, these relationships were consistent across age, sex, treatment group (different dosages of pregabalin/placebo), the five clinical conditions, and whether study results demonstrated separation from placebo or not (Farrar et al., 2001). To permit comparisons with previous studies and meta-analyses, investigators may also wish to report the percentages of patients obtaining reductions in pain intensity from baseline of at least 50% (McQuay and Moore, 1998).

3.1.2. Rescue analgesics and concomitant pain treatments

The use of all pain-related treatments during the course of a clinical trial should be assessed, including rescue analgesics and any other concomitant pain treatments. This is a straightforward task in single-dose analgesic trials that prohibit the concurrent use of other medications, but it is more difficult in chronic pain clinical trials that allow concurrent use of pain medications and other treatments for pain (e.g. physical therapy) for weeks or months. Some chronic pain trials have allowed previously used pain medications to be continued throughout the trial, and dosage stabilization is often required before patients are allowed to enroll in such trials. However, when changes in the use of concomitant pain treatments are permitted, they can be considered as an outcome measure (e.g. Kieburtz et al., 1998).

Providing patients with access to rescue analgesics may make it easier to include a placebo group in treatment efficacy studies, since patients not obtaining adequate pain relief are provided with an analgesic. However, administration of rescue treatment complicates the interpretation of differences in pain ratings between patients taking placebo and active treatments because of the reduction in pain expected to occur in patients receiving rescue treatment. The use of rescue medications is affected by both patient and provider beliefs. Patients use rescue medications to achieve varying levels of pain relief, but also for other reasons, including improving sleep or reducing anxiety, preventing increased pain resulting from increased activity, and treating pain (e.g. headache) that may be unrelated to the clinical trial. When recording treatments used for pain during the clinical trial, it may therefore be desirable to distinguish analgesics used for relief of the disorder being studied from all other uses.

Rescue medication consumption has been used as an outcome measure in clinical trials, with assessments including amount used and time-to-use (e.g. Chrubasik et al., 2003; Eisenberg et al., 2001). Scales have been developed that allow quantification of medication use in chronic pain patients based on dosage and medication class (Steedman et al., 1992), and composite measures have been proposed that combine rescue medication usage and pain intensity ratings into a single score (Lehmann, 1990; Silverman et al., 1993). Although these may be used to compare different treatment groups in clinical trials, the psychometric properties of such composite measures are not well established.

Despite the complex issues involved in the interpretation of rescue medication usage in a clinical trial, patients in a placebo group can be expected to take more of a rescue treatment than patients administered an efficacious investigational treatment. When considered together with pain intensity ratings, the amount of rescue treatment used by patients therefore can provide an important supplemental measure of the efficacy of the treatment being evaluated. For these reasons, assessments of rescue treatments are

recommended as a core outcome in trials where rescue interventions are available and permitted.

3.1.3. Pain quality and temporal aspects of pain

Pain has different sensory and affective qualities in addition to its intensity, and various measures of these components of pain can be used to more fully describe a patient's pain experience (Price et al., 1987). The efficacy of pain treatments may differ for various pain qualities. Measures of the affective and sensory qualities of pain may therefore identify treatments that are efficacious for certain aspects of pain but not for overall pain intensity. Assessment of pain qualities at baseline also makes it possible to determine whether certain patterns of pain characteristics moderate the effects of treatment.

Whereas pain intensity reflects the overall magnitude of the pain, pain affect can be viewed as reflecting the distress caused by the pain. Assessment of pain affect or unpleasantness is supported by the evidence that the affective component of pain can be empirically distinguished from pain intensity and may be differentially responsive to treatments (Jensen, 2003; Price, 1999). As with pain intensity, pain affect can be assessed with VAS, NRS, and VRS items having different anchors, for example, 'not unpleasant' and 'most unpleasant feeling possible'.

The Short-Form McGill Pain Questionnaire (SF-MPQ; Melzack, 1987) assesses 15 specific sensory and affective pain descriptors and provides a total score and sensory and affective subscale scores. This questionnaire is reliable and well-validated, and its sensory and affective subscales have demonstrated responsiveness in recent chronic pain clinical trials (e.g. Dworkin et al., 2003; Rowbotham et al., 1998). Because it assesses both specific sensory pain qualities and the affective component of pain, the SF-MPQ is recommended for inclusion in clinical trials as a secondary outcome measure to evaluate the effects of pain treatment on both sensory and affective qualities of pain.

Measures of the temporal aspects of pain—including variability in intensity; time to onset of meaningful pain relief; durability of pain relief; and frequency, duration, and intensity of pain episodes—have not received adequate attention in pain research. The available evidence indicates that measures of pain frequency have validity and represent a distinct dimension of pain (Jensen and Karoly, 2001). Frequency of 'breakthrough' pain (periods of severe pain superimposed on ongoing pain) is an important temporal aspect of pain that has been used as an outcome measure in clinical trials (e.g. Farrar et al., 1998). When appropriate, investigators should consider administering measures of the temporal aspects of pain as secondary outcome measures in clinical trials. The temporal dimensions that should be considered include patients' reports of the time to onset of meaningful pain relief and its durability as well as the frequency and intensity of episodes of breakthrough pain.

3.2. Physical functioning

Chronic pain interferes with daily activities, and it has been assumed that relief of pain is accompanied by improvement in function. However, many studies have demonstrated that pain intensity and physical functioning are only modestly associated (Turk, 2002), which supports the importance of including measures of functioning in chronic pain clinical trials. Measures of physical functioning typically assess multiple aspects of function, including activities of daily living. Disturbed sleep is prevalent in people with chronic pain, and its assessment is also important in chronic pain trials. Individuals with chronic pain consider both increased ability to function and improved sleep important treatment objectives (Casarett et al., 2001).

There are two broad types of measures of physical functioning and, more generally, health-related quality of life (HRQOL). Generic measures provide information about physical functioning and treatment benefits that can be compared across different conditions and studies; disease-specific measures assess problems associated with specific conditions that may not be assessed by generic measures and may also be more responsive to the effects of treatment (e.g. Dworkin et al., 2001; Guyatt et al., 1993). Because each of these approaches has strengths, use of both disease-specific measures, when available, and generic measures of physical functioning should be considered in designing chronic pain clinical trials.

On the basis of reviews of the literature on generic and pain-related measures of physical functioning prepared for the IMMPACT-II consensus meeting (Haythornthwaite, 2003; Stucki and Cieza, 2003) and discussions among the participants, use of a disease-specific measure of physical functioning is recommended in chronic pain clinical trials when a suitable and well-accepted one is available. Examples of such disease-specific measures of physical functioning are the WOMAC (Bellamy et al., 1988) and the Roland and Morris Back Pain Disability Scale (Roland and Morris, 1983). However, disease-specific measures of physical functioning have not been developed and validated for many chronic pain conditions. In clinical trials examining such disorders, use of either the Multidimensional Pain Inventory (MPI; Kerns et al., 1985) Interference Scale or the Brief Pain Inventory (BPI; Cleeland and Ryan, 1994; Cleeland et al., 1996) pain interference items (i.e. general activity, mood, walking ability, work, relations with other people, sleep, enjoyment of life) is recommended. The MPI and BPI interference scales both provide reliable and valid measures of the interference of pain with physical functioning that have been translated into many languages and studied in diverse chronic pain conditions in multiple countries.

The MPI and BPI measures of physical functioning have distinct advantages and disadvantages, and use of both may be considered when doing so would not impose an undue

burden on participants (a total of 16 items, 9 for the MPI and 7 for the BPI). The MPI Interference Scale does not assess sleep, and if this measure of physical functioning is administered, then use of a reliable and valid measure of the impact of pain on sleep is recommended. The BPI does include an item assessing pain interference with sleep, but also includes ratings of mood, social relations, and enjoyment of life. These three items may constitute a separate factor measuring affective state that is relatively independent of the remaining items (Cleeland et al., 1996). Few clinical trials, however, have examined BPI factors separately and so administration and analysis of only the three BPI activity items (general activity, walking ability, normal work) as a measure of physical functioning cannot be recommended until more data become available.

Regardless of whether a disease-specific measure of physical functioning or the MPI or BPI interference scale is used in a clinical trial, administration of a generic measure of physical functioning should be considered to obtain data that will allow comparisons with other disorders and that could be used in cost-effectiveness analyses (Thompson, 2002; Turk, 2002). The SF-36 Health Survey (Ware and Sherbourne, 1992) is the most commonly used generic measure of HRQOL and it has been used in studies of diverse medical and psychiatric disorders and in numerous clinical trials. The authors recommend the SF-36 as a generic measure of physical functioning because of the large amount of data available that permit comparisons among different disorders and treatments. The development of new HRQOL measures is an active area of research and these may offer improvements over the SF-36 and ultimately replace it (e.g. Chwastiak and Von Korff, 2003; Rogers et al., 2000).

In many chronic pain conditions, increased activity is accompanied by increased pain. Some patients limit their physical functioning because of pain, and their response to decreased pain may be to increase their activity until pain increases to its tolerated intensity. Other patients will tolerate increased pain to maintain a desired level of function and their response to decreased pain may be to report less pain as long as their level of function remains satisfactory. Although both situations represent true relief of pain, pain relief is reflected in increased activity with little change in pain intensity in the first, and in decreased pain intensity with little change in activity in the second. This issue has been addressed in some studies by examining combined measures of activity level and pain intensity to assess outcome (Malec et al., 1981; Peters and Large, 1990), but additional research on such composite measures is needed.

3.3. Emotional functioning

Chronic pain is often accompanied by symptoms of psychological distress and psychiatric disorders, including depression, anxiety, and anger (Fernandez, 2002). On the basis of a review of the literature of measures of emotional functioning prepared for the IMMPACT-II consensus meeting (Kerns, 2003) and discussions among the participants, the Beck Depression Inventory (BDI; Beck et al., 1961) and the Profile of Mood States (POMS; McNair et al., 1971) are recommended as core outcome measures of emotional functioning in chronic pain clinical trials. Both the BDI and POMS have well-established reliability and validity in the assessment of symptoms of depression and emotional distress, and they have been used in numerous clinical trials in psychiatry and in an increasing number of chronic pain clinical trials (Kerns, 2003). In research in psychiatry and in chronic pain, the BDI provides a well-accepted measure of the level of depressed mood in a sample and its response to treatment.

The POMS assesses six mood states—tension–anxiety, depression–dejection, anger–hostility, vigor–activity, fatigue–inertia, and confusion–bewilderment—and also provides a summary measure of total mood disturbance. Although the discriminant validity of the POMS scales in patients with chronic pain has not been adequately documented, the POMS has scales for the three most important dimensions of emotional functioning in chronic pain patients (depression, anxiety, anger) and also assesses three other dimensions that are very relevant to chronic pain and its treatment, including a positive mood scale of vigor–activity. Moreover, the POMS has demonstrated beneficial effects of treatment in some (but not all) recent chronic pain trials (e.g. Rowbotham et al., 1998). For these reasons, administration of both the BDI and the POMS is recommended in chronic pain clinical trials to assess the major aspects of the emotional functioning outcome domain.

The assessment of emotional functioning in patients with chronic pain presents a challenge because various symptoms of depression—such as decreased libido, appetite or weight changes, fatigue, and memory and concentration deficits—are also commonly believed to be consequences of chronic pain and the medications used for its treatment (Gallagher and Verma, 2004). It is unclear whether the presence of such symptoms in patients with chronic pain (and other medical disorders) should nevertheless be considered evidence of depressed mood, or whether the assessment of mood in these patients should emphasize symptoms that are less likely to be secondary to physical disorders (Wilson et al., 2001). Because the evidence indicates that measures of emotional functioning are adequately reliable, valid, and responsive when used in the medically ill (Kerns, 2003), the authors recommend that the principal analyses of the BDI and POMS in chronic pain clinical trials use the original versions without adjustment for presumed confounding by somatic symptoms. Depending on the specific objective of the clinical trial, supplemental analyses could be conducted to separately examine non-somatic and somatic aspects of emotional functioning.

3.4. Participant ratings of global improvement and satisfaction with treatment

Global ratings of improvement and satisfaction in a clinical trial provide an opportunity for participants to aggregate all of the components of their experience—pain relief, improvement in physical and emotional functioning, side effects, convenience—into one overall measure of their perception of the advantages and disadvantages of the treatment they received. Such measures reflect the 'disdisparate values and preferences of individual patients' (Gill and Feinstein, 1994) and in so doing provide an important measure of pain treatment outcome (Collins et al., 2001). Moreover, global ratings by patients of their improvement and satisfaction with treatment can be used to investigate participants' judgments of the clinical importance of changes in other outcome measures (Farrar et al., 2001; Fischer et al., 1999).

Many different approaches have been used to assess participants' overall evaluation of their treatment in clinical trials. On the basis of a review of the literature of measures of global outcome prepared for the IMMPACT-II consensus meeting (Farrar, 2003) and discussions among the participants, the Patient Global Impression of Change scale (PGIC; Guy, 1976) is recommended for use in chronic pain clinical trials as a core outcome measure of global improvement with treatment. This measure is a single-item rating by participants of their improvement with treatment during a clinical trial on a 7-point scale that ranges from 'very much improved' to 'very much worse' with 'no change' as the mid-point.

There has been widespread use of the PGIC in recent chronic pain clinical trials (e.g. Dunkl et al., 2000; Farrar et al., 2001), and the data provide a responsive and readily interpretable measure of participants' assessments of the clinical importance of their improvement or worsening over the course of a clinical trial. Impression of change scores using different verbal outcome categories have also been used to determine the minimally important changes in quality of life measures (e.g. Guyatt et al., 2002; Hägg et al., 2003). Although these measures appear to have validity, additional research is necessary to determine the relative extent to which ratings on the PGIC and similar measures reflect reduced pain, improvement in functioning, side effect burden, or other variables and whether this varies for different samples and treatments.

Other approaches to the global assessment of treatment response that have been used less frequently than the PGIC in chronic pain trials include ratings of participant satisfaction with treatment, prospectively conducted global ratings of disease state from which changes from baseline can be calculated, and global ratings of specific outcome domains, for example, global ratings of improvement in physical functioning or in overall side-effect burden (Middell et al., 2001). Single-item ratings of treatment outcome have both advantages and disadvantages when compared to multiple-item scales (Sloan et al., 2002), and additional research will be important to determine the optimal method for obtaining global ratings from patients.

3.5. Symptoms and adverse events

The assessment, analysis, and reporting of adverse events is an essential component of all clinical trials. Within the context of pharmacologic investigations, adverse events have been defined as "any untoward medical occurrence in a patient or clinical investigation participant administered a pharmaceutical product and which does not necessarily have to have a causal relationship with this treatment" (International Conference on Harmonisation, 1995b, p. 2–3). Such events are unintended signs, symptoms, laboratory abnormalities, or diseases associated in a temporal manner with the use of a medication.

Clinical trial protocols should define the method of assessment and the rationale for that approach. In selecting the approach used for ascertaining adverse events and the methods used for recording and coding the terms used to describe these events (e.g. Medical Dictionary for Regulatory Activities, Brown, 2003), consideration should be given to the type and purpose of the trial, whether international regulatory requirements dictate certain approaches (International Conference on Harmonisation, 1995a,b, 1996b), the phase of development or post-marketing, and the total safety experience with the product.

On the basis of a review of the literature on the assessment of symptoms and adverse events prepared for the IMMPACT-II consensus meeting (Katz, 2003) and discussions among the participants, the authors recommend that, at a minimum, passive capture of spontaneously reported events and the use of open-ended prompts should be used in chronic pain clinical trials to assess adverse events. In describing the results of clinical trials, the incidence of individual adverse events and serious adverse events should be reported for each treatment group, including the percentages of participants who experienced treatment emergent adverse events of particular significance or with an incidence greater than placebo. It is also very important to evaluate and report the severity of adverse events as this may differ among treatments that have a comparable incidence of adverse events (Edwards et al., 1999).

Active capture using structured interviews or questionnaires to assess specific symptoms and adverse events that are relevant to the disorder or treatment being studied will often be more sensitive and more informative than passive capture or general inquiries (e.g. Anderson and Testa, 1994; Edwards et al., 1999). Depending on the objectives of a chronic pain clinical trial, active capture of selected symptoms and adverse events can be conducted at periodic intervals throughout the trial, including baseline and the conclusion of the trial, ideally by the same investigator.

It is important to recognize that the frequency, duration, intensity, distress, importance to the patient, impact on daily

function, and investigator and patient causal attributions can be assessed for symptoms and adverse events (e.g. Anderson et al., 1999; Portenoy et al., 1994; Wolfe et al., 2000). Such assessments provide information about the clinical importance of safety and tolerability outcomes.

The authors recommend that methods for active capture of symptoms and adverse events relevant to chronic pain and its treatment be vigorously explored. In developing comprehensive strategies to assess these events, consideration should be given to including participant ratings of frequency, severity, importance, and associated distress. In such research, it will be important to evaluate whether the use of these methods increases the reported incidence of clinically insignificant events that have no implications for tolerability, safety, and patient satisfaction with treatment.

3.6. Participant disposition

On the basis of a review of the literature on the assessment of participant disposition in clinical trials prepared for the IMMPACT-II consensus meeting (Turk, 2003) and discussions among the participants, the authors recommend that chronic pain clinical trials should collect and report comprehensive information on participant disposition, including detailed information regarding the recruitment of participants and their progression through the trial. Information on participant disposition is essential for the adequate evaluation of the results of a clinical trial and for interpreting the trial's conclusions regarding efficacy and safety.

Although the CONSORT (Consolidated Standards of Reporting Trials) guidelines (Altman et al., 2001; Begg et al., 1996; Moher et al., 2001) were developed to serve as a guide to reporting results of clinical trials, they also provide a valuable enumeration of the core elements of information on participant disposition that should be recorded when conducting trials (Goudas et al., 2001), including the numbers of participants who withdraw and are lost to follow-up as well as the reasons for withdrawal and loss. The following additional information can be valuable in interpreting the results of a clinical trial and should be collected and reported when doing so is feasible: (1) the recruitment process and the percentages of participants enrolled from each recruitment method; (2) the number of candidate participants who were excluded from participation and the reasons why; (3) the number of candidates who chose not to enter the trial and the reasons why; (4) the use of prohibited concomitant medications and all other protocol deviations that may impact the interpretation of the trial results; (5) the number and reasons for withdrawal from each treatment group, including deaths and patients lost to follow up; and (6) the types, rates, and reasons for nonadherence with treatment in each treatment group.

Dosages and duration of all treatments received by participants during the clinical trial should be recorded, including assessments of the use of rescue, concomitant, and prohibited medications and all alterations in prescribed treatment. Detailed information describing the extent to which each participant adhered to the protocol will make it possible for data analyses to be conducted that specifically examine efficacy in patients who adhered to the protocol. Such *efficacy evaluable* or *per protocol* analyses can sometimes be valuable in interpreting the results of intention-to-treat analyses, although the benefits of comparing randomized groups are lost. Although an important component of patient disposition, withdrawal from a clinical trial due to lack of treatment effectiveness can also be considered an endpoint (European Agency for the Evaluation of Medicinal Products, 2002; International Conference on Harmonisation, 2001).

Although reasons for withdrawal are usually provided in reports of clinical trials, this information is often inadequate. For example, 'drop out due to adverse events' may be given as a reason for withdrawal, but this is not informative without tabulation of the specific adverse events associated with the withdrawals. Similarly, 'withdrawal of consent' is commonly given as a reason for withdrawals, but this is impossible to interpret without description of the reasons why patients withdrew consent.

There are several factors that may compromise the integrity of the double-blind used in a clinical trial (Even et al., 2000). Participants' and investigators' guesses of which treatment was administered should therefore be assessed, and the reasons for the specific guesses (e.g. medication side effects or pain relief) should also be collected to assist in intepreting any unblinding that may have occurred (Moscucci et al., 1987; Turner et al., 2002).

4. Conclusions

The authors recommend that the core outcome measures listed in the table should be *considered* when designing clinical trials of chronic pain treatments. It must be emphasized, however, that the authors are not suggesting that the inclusion of these measures in a trial should be considered a requirement for publication in a scientific journal or by regulatory agencies. Furthermore, these recommendations are not meant to imply that positive results must be obtained for all of these outcome measures for a treatment to be deemed efficacious and safe.

Pain intensity and impairments in physical and emotional functioning are associated in patients with chronic pain, and improvement in pain has been associated with improvement in functioning and reports of overall benefit in some but not all clinical trials. There are many circumstances, however, in which improvement is found for measures of one or two of the core outcome domains but not others. There are undoubtedly many explanations for such results, including the generally modest relationships among the core outcome domains (Turk et al., 2003). Moreover, the statistical power of clinical trials is typically determined for the primary

endpoint, and it can therefore be expected that inadequate power may sometimes explain results for secondary outcome measures that are not statistically significant. Conversely, positive results for secondary outcomes in chronic pain trials such as physical and emotional functioning would not necessarily provide convincing evidence of efficacy or adequate demonstration within regulatory contexts to support additional efficacy claims.

It is important to emphasize that there will be clinical conditions or treatments for which one or more of these core outcome measures will not be relevant and should therefore not be included in a clinical trial. Future research may also identify other measures of these core domains that will be shown to have psychometric properties that are superior to the specific measures recommended in this article. The authors also expect that the recommended measures will typically be supplemented by other measures that are included for exploratory purposes or to evaluate treatment- or disease-specific issues (Turk and Melzack, 2001). Regardless of which measures are ultimately used, the reasons for selecting each of the specific measures that have been included in a clinical trial should be provided.

There are many decisions that must be made in administering outcome measures in chronic pain trials. For example, whether ratings of pain or the other measures discussed in this article are made using retrospective or serial assessments is a very important issue that may have implications for the ability of a measure to detect change (Fischer et al., 1999). These and other decisions will depend on the design of the trial, the resources available, and other considerations that are beyond the scope of this article.

In recommending specific core outcome measures, the authors acknowledge the important limitations of existing measures and the pressing need to develop improved methods for assessing chronic pain outcomes. For this reason, forthcoming IMMPACT recommendations will provide guidelines for developing improved measures of chronic pain outcomes and will identify the types of studies that are required to successfully develop such measures. Additional IMMPACT meetings will focus on methods to identify the clinical importance of changes in chronic pain outcome measures, and on approaches for combining multiple outcome measures to evaluate treatment efficacy and effectiveness. The use of standard outcome assessments has the potential to greatly enhance the validity, comparability, and clinical applicability of clinical trials of chronic pain treatments.

5. Disclaimer

The views expressed in this article are those of the authors. No official endorsement by the US Department of Veterans Affairs, US Food and Drug Administration, US National Institutes of Health, or the pharmaceutical companies that provided unrestricted grants to the University of Rochester Office of Professional Education should be inferred.

Acknowledgements

The authors thank Charles S. Cleeland, PhD, for his comments on a preliminary draft of this article, and Paul J. Lambiase and Mary Gleichauf for their invaluable assistance. Abbott Laboratories, AstraZeneca, Elan Pharmaceuticals, Eli Lilly & Co., Endo Pharmaceuticals Inc., GlaxoSmithKline, Johnson and Johnson, Merck & Co., NeurogesX, Novartis Pharmaceuticals, Pfizer, and Purdue Pharma provided unrestricted grants to the University of Rochester Office of Professional Education to support the consensus meeting.

References

Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, Gøtzsche PC, Lang T, CONSORT Group. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med 2001;134:663–94.

Anderson RB, Testa MA. Symptom distress checklists as a component of quality of life measurement: comparing prompted reports by patient and physician with concurrent adverse event reports via the physician. Drug Inf J 1994;28:89–114.

Anderson RB, Hollenberg NK, Williams GH. Physical Symptoms Distress Index: a sensitive tool to evaluate the impact of pharmacological agents. Arch Intern Med 1999;159:693–700.

Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. Arch Gen Psychiatry 1961;4:561–71.

Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz KF, Simel D, Stroup DF. Improving the quality of reporting of randomized controlled trials: the CONSORT statement. J Am Med Assoc 1996;276:637–9.

Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988;15:1833–40.

Bellamy N, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P, Altman R, Brandt K, Dougados M, LeQuesne M. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis: consensus development at OMERACT III. J Rheumatol 1997;24:799–802.

Brown EG. Methods and pitfalls in searching drug safety databases utilising the Medical Dictionary for Regulatory Activities (MedDRA). Drug Safety 2003;26:145–58.

Casarett D, Karlawish J, Sankar P, Hirschman K, Asch DA. Designing pain research from the patient's perspective: what trial end points are important to patients with chronic pain? Pain Med 2001;2:309–16.

Chrubasik S, Model A, Black A, Pollak S. A randomized double-blind pilot study comparing Doloteffin® and Vioxx® in the treatment of low back pain. Rheumatology 2003;42:141–8.

Chwastiak LA, Von Korff M. Disability in depression and back pain: evaluation of the World Health Organization Disability Assessment Schedule (WHO DAS II) in a primary care setting. J Clin Epidemiol 2003;56:507–14.

Cleeland CS, Ryan KM. Pain assessment: global use of the Brief Pain Inventory. Ann Acad Med 1994;23:129–38.

Cleeland CS, Nakamura Y, Mendoza TR, Edwards KR, Douglas J, Serlina RC. Dimensions of the impact of cancer pain in a four country sample: new information from multidimensional scaling. Pain 1996;67: 267–73.

Collins SL, Edwards J, Moore RA, Smith LA, McQuay HJ. Seeking a simple measure of analgesia for mega-trials: is a single global assessment good enough? Pain. 2001;91:189–94.

Cruccu G, Anand P, Attal N, Garcia-Larrea L, Haanpää M, Jørum E, Serra J, Jensen TS. EFNS guidelines for neuropathic pain assessment. Eur J Neurol 2004;11:153–62.

Deyo RA, Battie M, Beurskens AHHM, Bombardier C, Croft P, Koes B, Malmivaara A, Roland M, Von Korff M, Waddell G. Outcome measures for low back pain research: a proposal for standardized use. Spine 1998;23:2003–13.

Dunkl PR, Taylor AF, McConnell G, Alfano AP, Conaway MR. Responsiveness of fibromyalgia clinical trial outcome measures. J Rheumatol 2000;27:2683–91.

Dworkin RH, Nagasako EM, Hetzel RD, Farrar JT. Assessment of pain and pain-related quality of life in clinical trials. In: Turk DC, Melzack R, editors. Handbook of pain assessment. 2nd ed. New York: Guilford Press; 2001. p. 659–92.

Dworkin RH, Corbin AE, Young JP, Sharma U, LaMoreaux L, Bockbrader H, Garofalo EA, Poole RM. Pregabalin for the treatment of postherpetic neuralgia: a randomized, placebo-controlled trial. Neurology 2003;60:1274–83.

Edwards JE, McQuay HJ, Moore RA, Collins SL. Reporting of adverse effects in clinical trials should be improved: lessons from acute postoperative pain. J Pain Symptom Manage 1999;18:427–37.

Eisenberg E, Lurie Y, Daoud D, Ishay A. Lamotrigine reduces painful diabetic neuropathy: a randomized controlled study. Neurology 2001; 57:505–9.

European Agency for the Evaluation of Medicinal Products. Note for guidance on clinical investigation of medicinal products for treatment of nociceptive pain; 2002 [www.emea.eu.int/pdfs/human/ewp/061200en.pdf].

Even C, Siobud-Dorocant E, Dardennes RM. Critical approach to antidepressant trials: blindness protection is necessary, feasible and measurable. Br J Psychiatry 2000;177:47–51.

Farrar JT. The global assessment of pain and related symptoms. Presented at the second meeting of the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT-II); April 2003 [www.immpact.org/meetings.html].

Farrar JT, Cleary J, Rauck R, Busch M, Nordbrock E. Oral transmucosal fentanyl citrate: randomized, double-blinded, placebo-controlled trial for treatment of breakthrough pain in cancer patients. J Natl Cancer Inst 1998;90:611–6.

Farrar JT, Young JP, LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain 2001;94:149–58.

Fernandez E. Anxiety, depression, and anger in pain: research findings and clinical options. Texas: Advanced Psychological Resources; 2002.

Fischer D, Stewart AL, Bloch DA, Lorig K, Laurent D, Holman H. Capturing the patient's view of change as a clinical outcome measure. J Am Med Assoc 1999;282:1157–62.

Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. Health Technol Assess 1998;2(14).

Gallagher RM, Verma S. Mood and anxiety disorders in chronic pain. In: Dworkin RH, Breitbart WS, editors. Psychosocial aspects of pain: a handbook for health care providers. Seattle: IASP Press; 2004. p. 589–606.

Gill TM, Feinstein AR. A critical appraisal of the quality of quality-of-life measurements. J Am Med Assoc 1994;272:619–26.

Goudas LC, Carr DB, Bloch R, Balk E, Ioannidis JPA, Terrin N, Gialeli-Goudas M, Chew P, Lau J. Management of cancer pain, vol. 1–2. Rockville, MD: Agency for Healthcare Research and Quality; 2001.

Guy W. ECDEU assessment manual for psychopharmacology (DHEW Publication No. ADM 76 338). Washington, DC: US Government Printing Office; 1976.

Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. J Chron Dis 1987;40:171–8.

Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. Ann Intern Med 1993;118:622–9.

Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR, Clinical Significance Consensus Meeting Group. Methods to explain the clinical significance of health status measures. Mayo Clin Proc 2002;77: 371–83.

Hägg O, Fritzell P, Nordwall A. The clinical importance of changes in outcome scores after teatment for chronic low back pain. Eur Spine J 2003;121:12–20.

Hays RD, Hadorn D. Responsiveness to change: an aspect of validity, not a separate dimension. Qual Life Res 1992;1:73–5.

Haythornthwaite JA. The assessment of pain-related physical function for clinical trials in chronic pain. Presented at the second meeting of the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT-II); April 2003 [www.immpact.org/meetings.html].

International Conference on Harmonisation. Guideline for industry, E1A: the extent of population exposure to assess clinical safety: for drugs intended for long-term treatment of non-life threatening conditions; 1995a [www.fda.gov/cder/guidance/index.htm].

International Conference on Harmonisation. Guideline for industry, E2A: clinical safety data management: definitions and standards for expedited reporting; 1995b [www.fda.gov/cder/guidance/index.htm].

International Conference on Harmonisation. Guideline for industry, E6: good clinical practice: consolidated guidance; 1996a [www.fda.gov/cder/guidance/index.htm].

International Conference on Harmonisation. Guideline for industry, E3: structure and content of clinical study reports; 1996b [www.fda.gov/cder/guidance/index.htm].

International Conference on Harmonisation. Guidance for industry, E10: choice of control group and related issues in clinical trials; 2001 [www.fda.gov/cder/guidance/index.htm].

Jadad AR. Randomised controlled trials: a user's guide BMJ Books, London; 1998 http://www.bmjpg.com/rct/contents.html.

Jadad AR, Cepeda MS. Ten challenges at the Intersection of clinical research, evidence-based medicine, and pain relief. Ann Emerg Med 2000;36:247–52.

Jensen MP. The validity and reliability of pain measures for use in clinical trials in adults. Presented at the second meeting of the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT-II); April 2003 [www.immpact.org/meetings.html].

Jensen MP, Karoly P. Self-report scales and procedures for assessing pain in adults. In: Turk DC, Melzack R, editors. New York: Guilford Press, 2001. p. 15–34.

Katz N. The measurement of symptoms and side effects in clinical trials of chronic pain. Presented at the second meeting of the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT-II); April 2003 [www.immpact.org/meetings.html].

Keefe FJ, Williams DA, Smith SJ. Assessment of pain behaviors. In: Turk DC, Melzack R, editors.. New York: Guilford Press; 2001. p. 170–87.

Kerns RD. Assessment of emotional functioning in pain treatment outcome research. Presented at the second meeting of the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT-II); April 2003 [www.immpact.org/meetings.html].

Kerns RD, Turk DC, Rudy TE. The West Haven-Yale Multidimensional Pain Inventory (WHYMPI). Pain 1985;23:345–56.

Kieburtz K, Simpson D, Yiannoutsos C, Max MB, Hall CD, Ellis RJ, Marra CM, McKendall R, Singer E, Dal Pan GJ, Clifford DB, Tucker T, Cohen B, AIDS Clinical Trial Group 242 Protocol Team. A randomized trial of amitriptyline and mexiletine for painful neuropathy in HIV infection. Neurology 1998;51:1682–8.

Lehmann KA. Patient-controlled analgesia for postoperative pain. In: Benedetti C, Chapman CR, Giron G, editors. Opioid analgesia: recent advances in systemic administration. New York: Raven Press; 1990. p. 297–334.

Malec J, Cayner JJ, Harvey RF, Timming RC. Pain management: long-term follow-up of an inpatient program. Arch Phys Med Rehabil 1981;62: 369–72.

McNair DM, Lorr M, Droppleman LF. Profile of Mood States. San Diego, CA: Educational and Industrial Testing Service; 1971.

McQuay H, Moore A. An evidence-based resource for pain relief. Oxford, UK: Oxford University; 1998.

Melzack R. The short-form McGill Pain Questionnaire. Pain 1987;30: 191–7.

Middel B, Stewart R, Bouma J, van Sonderen E, van den Heuvel WJA. How to validate clinically important change in health-related functional status: is the magnitude of the effect size consistently related to magnitude of change as indicated by a global question rating? J Eval Clin Pract 2001;7:399–410.

Moher D, Schulz KF, Altman DG, CONSORT Group. The CONSORT Statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. Lancet 2001;357:1191–4.

Moscucci M, Byrne L, Weintraub M, Cox C. Blinding, unblinding, and the placebo effect: an analysis of patients' guesses of treatment assignment in a double-blind clinical trial. Clin Pharmacol Ther 1987;41:259–65.

Peters JL, Large RG. A randomised control trial evaluating in- and outpatient pain management programmes. Pain 1990;41:283–93.

Portenoy RK, Thaler HT, Kornblith AB, Lepore JM, Friedlander-Klar H, Kiyasu E, Sobel K, Coyle N, Kemeny N, Norton L, Sher H. The Memorial Symptom Assessment Scale: an instrument for the evaluation of symptom prevalence, characteristics, and distress. Eur J Cancer 1994;30A:1326–36.

Price DD. Psychological mechanisms of pain and analgesia. Seattle: IASP; 1999.

Price DD, Harkins SW, Baker C. Sensory-affective relationships among different types of clinical and experimental pain. Pain 1987;28:297–307.

Rogers WH, Wittink H, Wagner A, Cynn D, Carr DB. Assessing individual outcomes during outpatient multidisciplinary chronic pain treatment by means of an augmented SF-36. Pain Med 2000;1:44–54.

Roland M, Morris R. A study of the natural history of back pain, part I: development of a reliable and sensitive measure of disability in low back pain. Spine 1983;8:141–4.

Rowbotham MC, Harden N, Stacey B, Bernstein P, Magnus-Miller L. Gabapentin Postherpetic Neuralgia Study Group: Gabapentin for the treatment postherpetic neuralgia: a randomized controlled trial. J Am Med Assoc 1998;280:1837–42.

Scientific Advisory Committee of the Medical Outcomes Trust. Assessing health status and quality-of-life instruments: attributes and review criteria. Qual Life Res 2002;11:193–205.

Silverman DG, O'Connor TZ, Brull SJ. Integrated assessment of pain scores and rescue morphine use during studies of analgesic efficacy. Anesth Analg 1993;77:168–70.

Sloan JA, Aaronson N, Cappelleri JC, Fairclough DL, Varricchio C, Clinical Significance Consensus Group. Assessing the clinical significance of single items relative to summated scores. Mayo Clin Proc 2002;77:479–87.

Steedman SM, Middaugh SJ, Kee WG, Carson DS, Harden RN, Miller MC. Chronic-pain medications: equivalence levels and method of quantifying usage. Clin J Pain 1992;8:204–14.

Stucki G, Cieza A. HRQL measures in chronic pain trials. Presented at the second meeting of the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT-II), April 2003 [www.immpact.org/meetings.html].

Terwee CB, Dekker FW, Wiersinga WM, Prummel MF, Bossuyt PMM. On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. Qual Life Res 2003;12:349–62.

Thompson D. Toward a pharmacoeconomic model of neuropathic pain. Clin J Pain 2002;18:366–72.

Turk DC. Clinical effectiveness and cost-effectiveness of treatments for patients with chronic pain. Clin J Pain 2002;18:355–65.

Turk DC. Reporting on participant disposition in clinical trials. Presented at the second meeting of the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT-II); April 2003 [www.immpact.org/meetings.html].

Turk DC, Melzack R. In: Turk DC, Melzack R, editors. Handbook of pain assessment. 2nd ed. New York: Guilford Press; 2001.

Turk DC, Dworkin RH, Allen RR, Bellamy N, Brandenburg N, Carr DB, Cleeland C, Dionne R, Farrar JT, Galer BS, Hewitt DJ, Jadad A, Katz NP, Kramer LD, Manning DC, McCormick CG, McDermott M, McGrath P, Quessy S, Rappaport BA, Robinson JP, Royal MA, Simon L, Stauffer JW, Stein W, Tollett J, Witter J. Core outcome domains for chronic pain clinical trials: IMMPACT recommendations. Pain 2003;106:337–45.

Turner JA, Jensen MP, Warms CA, Cardenas DD. Blinding effectiveness and association of pretreatment expectations with pain improvement in a double-blind randomized controlled trial. Pain 2002;99:91–9.

United States Department of Health and Human Services. Food and Drug Administration (Docket No. 97D-0188): International Conference on Harmonisation; guidance on general considerations for clinical trials Federal Register. 1997;62:66113–9www.fda.gov/cder/guidance/1857fnl.pdf.

Ware JE, Sherbourne CD. TheMOS36-item short-form health survey (SF-36). Med Care 1992;30:473–83.

Williams RC. Toward a set of reliable and valid measures for chronic pain assessment and outcome research. Pain 1988;35:239–51.

Wilson KG, Mikail SF, D'Eon JL, Minns JE. Alternative diagnostic criteria for major depressive disorder in patients with chronic pain. Pain 2001; 91:227–34.

Wolfe F, Kong SX, Watson DJ. Gastrointestinal symptoms and health related quality of life in patients with arthritis. J Rheumatol 2000;27: 1373–8.

Editorial

WOMAC: A 20-Year Experiential Review of a Patient-Centered Self-Reported Health Status Questionnaire

A key element in clinical research and clinical practice in musculoskeletal medicine is the evaluation of the therapeutic benefit of interventions used either singularly or in combination. In both clinical research and clinical practice environments, reliability, validity, and responsiveness are essential attributes of health status measurement tools, and in the latter brevity, simplicity, and ease of scoring are regarded with high importance[1,2].

Prior to 1981, measurement procedures for quantifying pain, stiffness, and physical disability in hip and knee osteoarthritis (OA) in rheumatology were diverse and lacked standardization in content, format, and scaling[3]. Further, health status questionnaires were available in very few languages, most often having been developed in English and translated into a few European languages.

The challenge in 1981 was to build a standardized disease-specific patient-relevant self-reported health status questionnaire for hip and knee OA. In 1982, I had the opportunity in the course of completing an MSc thesis to describe the development of a health status questionnaire termed the Western Ontario and McMaster (WOMAC) Osteoarthritis Index[4]. Twenty years later, the WOMAC Index has been extensively validated and has been translated and linguistically validated in over 60 alternative-language forms. In the majority of alternative-language forms it is available in both Likert (LK) and visual analog (VA) scaling formats. There are several hundred citations (full manuscripts, abstracts, reviews) to the use of WOMAC in validation studies, comparative studies against other health status measures, and in its application in various clinical research and clinical practice settings[5].

The idea for the WOMAC index evolved from a brief discussion with Professor Watson Buchanan, a conversation in which I sought his advice in selecting a thesis topic that would address an unmet need in clinical measurement. While development of the idea took only 12 months, the validation and implementation was to consume much of the next 15 years. Between 1996 and 1999 the Index underwent significant refinement, a process that has been consolidated between 1999 and the present, and has resulted in the 3.1 series of WOMAC questionnaires. The WOMAC LK3.1 and WOMAC VA3.1 versions of the Index are now extensively used, particularly in assessing efficacy in pharmaceutical and biotechnology environments.

The success of the WOMAC index is in large part related to 6 factors: (1) Extensive patient involvement in the development of the item inventory[6]. This is perhaps the most important since it is an approach that reduces the potential influence of paternalism, and anchors the item content into aspects of the disease experience that are relevant to patients, and to which they can therefore relate. (2) The conduct of numerous studies evaluating different clinimetric properties of the Index, including analyses evaluating validity, reliability, and responsiveness, comparative studies assessing LK versus VA scaling, blind versus informed presentation, tracking signal items versus complete index usage, parametric versus non-parametric analyses and time frame variations[5]. (3) The development and linguistic validation of numerous alternative-language forms of WOMAC VA3.1 and WOMAC LK3.1 using a standard operating procedure based on tandem forward and backward translation processes and subsequent linguistic validation[5]. (4) Continued research and development into content and administration issues including the application of WOMAC in telephone interviews[7] as well as mouse driven cursor and touch screen electronic data capture formats[8,9]. (5) The incorporation of WOMAC into Osteoarthritis Research Society International (OARSI) clinical trials guidelines as an index relevant to outcome measurement in OA[10]; and (6) the provision of the WOMAC Index, in the required scaling format, alternative-language form, and administration format for academic, commercial, and clinical applications, and ongoing user support.

The development of WOMAC has not been without its challenges. Trans-cultural adaptation of the WOMAC 3.1 Index has been a complex process for which Health Outcomes Group, Palo Alto, California, USA, have taken primary responsibility and in which they have applied their standard operating procedures to develop linguistically valid alternative-language forms of extremely high quality. The preponderance of instruments developed in either North America or Europe might be viewed with concern given the global nature of OA and diversity of lifestyles. It is gratifying, therefore, that the performance of the WOMAC Index has been maintained in its global applications. Thus, while potentially reflecting a restricted view of global diversity, the Index nevertheless appears to tap into the commonalities that exist in the

dimensionality of the symptomatology of OA. That notwithstanding, it is clear that the impact of environmental challenges involved in, for example, stair climbing and transportation are different in different parts of the world, and bathing and toileting habits are quite varied. The index constructor therefore is faced with the dilemma of whether to modify the item content and risk comparing apples with oranges or making minor accommodations in order to maintain a standard question battery. In the case of the WOMAC Index, the latter strategy has been followed, and allows a small degree of flexibility in interpretation. In terms of large multinational clinical trials I believe this is the preferred solution. However, it is possible that at a national level, for clinical practice applications, modification of the item inventory either at an individual question or group of questions (module) level might provide additional advantage.

We have assessed the performance of items self-selected by individual patients, a so-called signal strategy, but have been concerned by the inconsistency with which patients adhere to the selected signal with the passage of time. Our recommendation at the present time, therefore, is to use the entire Index, rather than the signal form.

Scaling format selection is a challenge for any instrument developer, trade-offs often being involved. Likert and VA scales are both commonly used in health status questionnaires. Likert scaling provides a simple and easing scoring system, while the more demanding VA scale may be slightly more sensitive. For this reason we have created parallel forms of the WOMAC 3.1, making available both LK and VA formats for most language forms. In the alternative-language forms, it has been interesting to note that even for the standard scales, word usage is different in different countries. For example, words such as "moderate" and "extreme" may be deemed appropriate in one context, but not in another. As a result the equivalent words may be "average" or "very severe," respectively, in some cultures.

I have been interested to note over the last several years that in some cases the WOMAC Index appears to have passed from one user to another and occasionally in that process the instrument has been altered in a variety of ways. Sometimes the modifications seem quite minor, such as crowding the questions on to one or 2 pages. On other occasions, more radical alterations of the Index have been made such as rescaling the instrument using Health Assessment Questionnaire-style scaling or using a 5 centimeter instead of 10 centimeter visual analog scale on a paper version of the instrument. From time to time I have been sent versions of the instrument that are incomplete, usually the result of the provider not having photocopied the entire instrument when passing it on to a friend or colleague. I have also encountered versions in which additional questions have been added but for which there is no apparent evidence of subsequent revalidation. The concern here is that some modifications may degrade instrument performance, or at the very least erode the level of standardization previously achieved. For this reason, and because the Index, even in English, exists in a number of different forms having different applications, I prefer to provide the most appropriate form of the Index directly to end users in order to better meet their specific measurement needs.

In comparative analyses against other disease-specific and generic health status measures, the WOMAC Index has frequently been superior in performance[11-14]. Two Rasch analyses using an item response theory approach to index construction seem to generally uphold the current structure, although this now popular approach might suggest some modification. However, the consequence of such modification on responsiveness has yet to be determined[15,16]. Recommendations both for shortening the Index[17] and for lengthening the Index[18] have been made, the former to reduce responder burden, the latter to encompass other, potentially younger and more athletic, individuals in orthopedic environments. A role for the WOMAC Index in predicting future health status[19] and health resource utilization[20] has been suggested, but remains to be clarified. Similarly, an application of the WOMAC Index in the assessment of lower limb involvement in rheumatoid arthritis has been suggested, but remains to be verified[21].

It is important to consider whether the development of the WOMAC Index is static or dynamic. The answer is most certainly that it is and remains distinctly dynamic. The developmental form of the WOMAC had 5 subscales (pain, stiffness, physical function, social function, emotional function), the first 3 of which were retained in the original form of the WOMAC and probed the symptom experience of OA in the "hips/and or knees." The WOMAC 3.0 focused on an investigator selected study joint. During that phase of development we also experimented with strategic variations such as using separate WOMAC indices for the study knee and the contralateral knee, and using separate WOMAC pain and stiffness subscales for the left and right knees but a common WOMAC physical function subscale. We have experimented with setting the time frame at 24 h, 48 h (WOMAC 3.1), past 7 days (WOMAC 3.1W), and past month (WOMAC 3.1M), and have created alternative-language forms and a signal version (WOMAC 3.1S). The development of the alternative-language translations has resulted in enhancements to the instructions to patients, the subscale introductory comments, the question stems, and to the WOMAC User Guide. We have looked at short-forming the Index (WOMAC 3.1SF), initial analyses suggesting the preferred short form may be in part dependent on clinical setting, geographic environment, and analytic strategy[22]. Opportunities for electronic data capture by computer-assisted technology have resulted in programs looking at alternatives to patient in-office self-completion on paper[9]. We are currently engaged in an initiative to assess the added value, from an effectiveness and cost-effectiveness standpoint, of providing quantitative WOMAC data to practitioners in a routine clinical care setting. We are also examining an expansion of the current WOMAC inventory

(WOMAC 3.1ER) to accommodate some potential opportunities that may exist in the study of purported structure-modifying OA drugs. WOMAC data have been used in developing a definition of minimum perceptible clinical improvement[23], and together with data from other instruments in developing the OARSI Responder Criteria[24]. We are further evaluating a weighting and aggregation system for the WOMAC Index using a device called the Patient Assessment of the Relative Importance of Symptoms (PARIS) Sectogram[5], and examining the relationship between WOMAC scores and scores from several generic health-related quality of life measures in patients with and without comorbidity. We are currently redeveloping the WOMAC website at www.womac.org to enhance information flow with new and established WOMAC users. An additional consequence of the WOMAC development has been the advantage provided by that experience, in the rapid development of a comparable index, termed the Australian/Canadian (AUSCAN 3.0) Index[25,26], for OA hand studies, details of which can be located at www.auscan.org. All these activities are indicative of a dynamic longterm commitment to advance and refine patient-centered outcome measurement in OA, for application in clinical research and clinical practice environments.

The last 20 years' development of the WOMAC has not been simply the application of classical measurement theory to symptom quantification. It has also involved an extensive collaboration with colleagues in musculoskeletal medicine and other health disciplines, and the interest and commitment of many patients with knee and/or hip OA. I am most grateful to all those who have given their time and resources to support this international initiative. The principal challenges now are to make a good measure even better, to maintain its relevancy in a changing multicultural world, to broaden its application in clinical practice environments, particularly considering issues such as individual response, shared goal setting, and personal and environmental modulators of outcome, to meet emerging needs in structure modifying environments, and to take advantage of emerging technological opportunities.

NICHOLAS BELLAMY, MD, MSc, MBA,
Professor and Director,
Centre of National Research on Disability
and Rehabilitation Medicine (CONROD),
The University of Queensland,
Mayne Medical School, Level 3,
Herston Road, Brisbane, Queensland 4006, Australia.

Address reprint requests to Prof. Bellamy.

REFERENCES

1. Bellamy N, Kaloni S, Pope J, Coulter K, Campbell J. Quantitative rheumatology: A survey of outcome measurement procedures in routine rheumatology outpatient practice in Canada. J Rheumatol 1998;25:852-8.
2. Bellamy N, Muirden KD, Brooks PM, Barraclough D, Tellus MM, Campbell J. Quantitative rheumatology: A survey of outcome measurement procedures in routine rheumatology outpatient practice in Australia. J Rheumatol 1999;26:1593-9.
3. Bellamy N, Buchanan WW. Outcome measurement in osteoarthritis clinical trials: The case for standardisation. Clin Rheumatol 1984;3:293-303.
4. Bellamy N. Osteoarthritis — an evaluative index for clinical trials [MSc thesis]. McMaster University, Hamilton, Ontario, Canada; 1982.
5. Bellamy N. WOMAC Osteoarthritis Index User Guide. Version V. Brisbane: Australia; 2002.
6. Bellamy N, Buchanan WW. A preliminary evaluation of the dimensionality and clinical importance of pain and disability in osteoarthritis of the hip and knee. Clin Rheumatol 1986;5:231-41.
7. Bellamy N, Campbell J, Hill J. A comparative study of telephone vs on-site completion of the WOMAC 3.0 Osteoarthritis Index. J Rheumatol 2002;29:783-6.
8. Bellamy N, Campbell J, Stevens J, Pilch L, Stewart C, Mahmood Z. Validation study of a computerized version of the Western Ontario and McMaster Universities VA3.0 Osteoarthritis Index. J Rheumatol 1997;24:2413-5.
9. Theiler R, Speilberger J, Bischoff HA, Bellamy N, Huber J, Kroesen S. Clinical evaluation of the WOMAC 3.0 OA Index in numeric rating scale format using a computerised touch screen version. Osteoarthritis Cart 2002;10:479-81.
10. Osteoarthritis Research Society (OARS) Task Force Report. Design and conduct of clinical trials of patients with osteoarthritis: recommendations from a task force of the Osteoarthritis Research Society. Osteoarthritis Cart 1996;4:217-43.
11. Hawker G, Melfi C, Paul J, Green R, Bombardier C. Comparison of a generic (SF-36) and a disease-specific (WOMAC) instrument in the measurement of outcomes after knee replacement surgery. J Rheumatol 1995;22:1193-6.
12. March LM, Oh E-S, Cross M, et al. A comparison of WOMAC and MOS SF-36 in OA patients undergoing joint replacement. 8th APLAR Congress of Rheumatology. April 21-26, 1996, Melbourne, Australia. Programme and Abstracts. Abstract #439.
13. Davies GM, Watson DJ, Bellamy N. Comparison of the responsiveness and relative effect size of the WOMAC and SF-36 in a randomized clinical trial in patients with osteoarthritis. Arthritis Care Res 1999;12:172-9.
14. Stucki G, Sangha O, Stucki S, et al. Comparison of the WOMAC (Western Ontario and McMaster Universities) Osteoarthritis Index and a self-report format of the self-administered Lequesne-Algofunctional Index in patients with knee and hip osteoarthritis. Osteoarthritis Cart 1998;6:79-86.
15. Wolfe F, Kong SX. Rasch analysis of the Western Ontario McMaster questionnaire (WOMAC) in 2205 patients with osteoarthritis, rheumatoid arthritis, and fibromyalgia. Ann Rheum Dis 1999;58:563-8.
16. Ryser L, Wright BD, Aeschlimann A, Mariacher-Gehler S, Stucki G. A new look at the Western Ontario and McMaster Universities Osteoarthritis Index using Rasch analysis. Arthritis Care Res 1999;12:331-5.
17. Whitehouse SL, Lingard EA, Learmonth ID, and Kinemax Outcomes Group. A reduced WOMAC function scale — derivation and validation [abstract]. Arthritis Rheum 2000;43 Suppl:S393.
18. Roos EM, Roos HP, Lohmander LS. WOMAC Osteoarthritis Index — additional dimensions for use in subjects with post-traumatic osteoarthritis of the knee. Western Ontario and McMaster Universities. Osteoarthritis Cart 1999;7:216-21.
19. Whitehouse SI, Lingard EA, Learmouth ID. Reduced WOMAC Function Scale validation and derivation [abstract]. Arthritis Rheum 2000;43 Suppl:S393.
20. Ethgen O, Kahler KH, Kong SX, Reginster Y-Y, Wolfe F. Are health-related quality of life scores useful in predicting the use of

health care services? Exploration in patients with arthritis [abstract]. Arthritis Rheum 2000;43 Suppl:S163.
21. Hobby KJ. The effects of aerobic walk-based exercise on women with rheumatoid arthritis [MSc thesis]. London: University of Western Ontario; 1994:1-239.
22. Bellamy N, Bolognese J, Barlas S, et al. Evaluation of factors influencing the item content of the Short Form WOMAC VA 3.1 (SFWOMAC VA3.1) Osteoarthritis Index [abstract]. Arthritis Rheum 2002;46 Suppl:S114.
23. Ehrich EW, Davies GM, Watson DJ, Bolognese JA, Seidenberg BC, Bellamy N. Minimal perceptible clinical improvement with the Western Ontario and McMaster Universities osteoarthritis index questionnaire and global assessments in patients with osteoarthritis. J Rheumatol 2000;27:2635-41.
24. Dougados M, LeClaire P, van der Heijde D, Bloch DA, Bellamy N, Altman RD. Special Article: Response criteria for clinical trials on osteoarthritis of the knee and hip: A report of the Osteoarthritis Research Society International Standing Committee for Clinical Trials Response Criteria Initiative. Osteoarthritis Cart 2000; 8:395-403.
25. Bellamy N, Campbell J, Haraoui B, et al. Dimensionality and clinical importance of pain and disability in hand osteoarthritis: development of the Australian/Canadian (AUSCAN) Osteoarthritis Hand Index. Osteoarthritis Cart 2002; (In Press).
26. Bellamy N, Campbell J, Haraoui B, et al. Clinimetric properties of the AUSCAN Osteoarthritis Hand Index: an evaluation of reliability, validity and responsiveness. Osteoarthritis Cart 2002; (In Press).

SECTION 7 - Flexible Delivery

The traditional method of administration of the WOMAC Index is by patient self-completion of the paper version. However, envisioning developments in electronic data capture (EDC) in research environments and applications of the WOMAC Index in clinical practice, population or community-based research studies, or decision-making in health care management, alternative delivery platforms for the WOMAC Index were explored.

Early work with a mouse-driven cursor application on a NeXT computer, and subsequently modified for use on a personal computer, was successful in developing the concept of EDC with the WOMAC Index, and validating an e-version of the WOMAC VA3.0 Index (27). The software was not completely debugged, and because of the emergence of more readily available touch-screen technology, we halted further development on this particular application. Nevertheless, the study was successful in demonstrating that valid WOMAC e-data, comparable to paper-based WOMAC data, could be acquired by patient self-completion at a computer terminal and transmitted to a remote location in another city (27).

In preparation for a pharmaco-economic evaluation of hylan G-F 20 in knee OA (17,18), a comparative study was undertaken comparing telephone versus onsite completion of the WOMAC LK3.0 Index (28). The purpose of the study was to explore whether valid WOMAC data could be obtained by a telephone interviewer, thus obviating the need for a patient to visit the doctor's office to self-complete the WOMAC LK3.0 Index (28). The study demonstrated that the method was both feasible and valid, and facilitated the delivery of the WOMAC Index by telephone in the aforementioned pharmaco-economic evaluation of hylan G-F 20 (17,18).

By late 1999, touch screen technology was evolving, but software was still not commercially available to support EDC of WOMAC data by touch-screen, and hardware was less than ideal for the application. A collaborative project, with colleagues in Switzerland, was initiated to build a touch-screen computer and write computer code for software to support patient self-completion of the e-WOMAC Index on a laptop computer (29). The study employed an 11-point numerical rating scale German for Switzerland version of the Index (WOMAC NRS 3.0) (29). The study demonstrated that the method was feasible, and e-WOMAC data comparable to paper-based WOMAC data were obtained, thus demonstrating the validity of this method of data capture using touch-screen technology. The responsiveness of this version of the e-WOMAC Index was not evaluated.

A further collaborative project, with colleagues in Switzerland, based on the German for Switzerland 5-point Likert version of the WOMAC Index was initiated, and utilised a commercially available touch-screen laptop computer and new purpose-built software to support patient self-completion of the e-WOMAC Index (30). The study demonstrated that the method was feasible, and e-WOMAC data comparable to paper-based WOMAC data were obtained, thus demonstrating the validity of this method of data capture using touch-screen technology (30). Furthermore, the responsiveness of the e-WOMAC LK 3.0 Index was evaluated in a longitudinal study, and found not to differ from that of the paper version (30).

Having worked with different versions of the WOMAC Index scaled on 5-point Likert, 11-point numerical and 100 mm visual analogue scaling formats, a study was undertaken to explore the relative responsiveness of different scaling formats (31). The study did not involve the WOMAC Index, but employed instead single global pain questions scaled in different formats, as well as the McGill Pain Questionnaire (MPQ). The patterns of joint involvement with OA varied between patients and were not confined to the hip or knee joints. The observations were consistent with our previous experience with the WOMAC Index, that the standardised response mean, a measure of responsiveness, was greater for the VA than the Likert scale. A similar analysis conducted by our group in patients with rheumatoid arthritis showed a similar scale-dependent pattern, as did a subsequent experience with a WOMAC-like Index, called the Australian/Canadian (AUSCAN) Hand Osteoarthritis Index, also originated by the author of this thesis. The 100 mm VA scale generally, therefore, appears to be more responsive than the 5-point Likert scaled equivalent (31).

Based on postal surveys conducted in the late 1990s in Canada (32) and Australia (33), responding clinical rheumatologists in both countries placed value for routine clinical care applications, not only on an assessment technique's validity, reliability and responsiveness, but also on its simplicity, quick completion and easy scoring. Accepting the limitation of the survey method, this requirement did not appear specific to OA, and likely represented a general trend, that might, in part, explain the relatively low level of uptake in routine clinical practice of health status questionnaires used relatively frequently in clinical research environments in studies of rheumatoid arthritis, osteoarthritis, ankylosing spondylitis and fibromyalgia (32,33).

In order to explore the opportunities for, and consequences of, shortening the item inventory of the WOMAC Index, a study was initiated, in collaboration with colleagues in France, on the longest of the WOMAC Index subscales, namely the physical function subscale (34). The investigation, using the WOMAC LK 3.0 Index in French for Canada, in 1218 outpatients, resulted in the cautious proposal of an 8-item short form of the normally 17-item long function subscale (34). The need for further studies, of different language translations and scaling formats, in different countries, clinical environments and interventions was acknowledged (34). Nevertheless, this study provided preliminary insight into opportunities for short-forming, in order to meet clinical practice user needs for simplicity and quick completion, in health status assessment questionnaires.

References:

27. Bellamy, N., Campbell, J., Stevens, J., Pilch, L., Stewart, C. and Mahmood, Z. Validation study of a computerized version of the Western Ontario and McMaster Universities VA3.0 Osteoarthritis Index. Journal of Rheumatology 1997;24:2413-2415.

28. Bellamy, N., Campbell,J. and Hill J. A comparative study of telephone vs on-site completion of the WOMAC 3.0 Osteoarthritis Index. Journal of Rheumatology 2002;29:783-786.

29. Theiler R., Speilberger J., Bischoff H.A., Bellamy N., Huber J. and Kroesen S. Clinical evaluation of the WOMAC 3.0 OA Index in numeric rating scale format using a touch screen version. Osteoarthritis and Cartilage 2002;10:479-481.

30. Theiler R., Bischoff-Ferrari, H.A., Good M. and Bellamy N. Responsiveness of the electronic touch screen WOMAC 3.1 OA Index in a short term clinical trial with rofecoxib. Osteoarthritis and Cartilage 2004;12:912-916.

31. Bellamy, N., Campbell, J. and Syrotuik, J. Comparative study of self-rating pain scales in osteoarthritis patients. Current Medical Research and Opinion 1999; 15(2):113-119.

32. Bellamy, N., Kaloni, S., Pope, J., Coulter, K., and Campbell, J. Quantitative rheumatology: A survey of outcome measurement procedures in routine rheumatology outpatient practice in Canada. The Journal of Rheumatology 1998;25:852-8.

33. Bellamy, N., Muirden, K.D., Brooks, P.M., Barraclough, D., Tellus, M.M., and Campbell, J. Quantitative rheumatology: A survey of outcome measurement procedures in routine rheumatology outpatient practice in Australia. The Journal of Rheumatology. 1999;26:1593-1599.

34. Tubach F., Baron G., Falissard B., Longeart I., Dougados M., Bellamy N. and Ravaud P. Using patients' and rheumatologists' opinions to specify a short form of the WOMAC functional subscale. Annals of the Rheumatic Diseases. 2005;64(1):75-79.

Validation Study of a Computerized Version of the Western Ontario and McMaster Universities VA3.0 Osteoarthritis Index

NICHOLAS BELLAMY, JANE CAMPBELL, JOHN STEVENS, LAWRENCE PILCH, CINDY STEWART, and ZAHID MAHMOOD

ABSTRACT. Objective. To study the validity and feasibility of a computerized version of the Western Ontario and McMaster Universities (WOMAC) Osteoarthritis Index.

Methods. Thirty patients with osteoarthritis (OA) of the knee completed both a paper and a computerized version of WOMAC in random order. The visual analog scaled version of WOMAC, VA3.0, was used. We studied criterion validity by comparing the paper and computerized versions.

Results. All patients completed the computerized version without undue difficulty. Criterion validity, based on aggregated subscale scores, was excellent: Pain, ICC = 0.89, Stiffness, ICC = 0.87, Physical Function, ICC = 0.95.

Conclusion. The computerized version of WOMAC VA3.0 is a valid alternative to the paper version. (*J Rheumatol 1997;24:2413–5*)

Key Indexing Terms:
WOMAC OSTEOARTHRITIS INDEX
OUTCOME MEASUREMENT
VALIDITY
OSTEOARTHRITIS CLINICAL TRIALS

Initially developed in 1982[1], the Western Ontario and McMaster Universities (WOMAC) Osteoarthritis Index has been the subject of 2 major validation studies[2,3]. The index takes the form of a self-administered questionnaire of 24 questions categorized in 3 subscales (Pain, Stiffness, Physical Function). In addition to verifying the adequacy of the reliability, validity (face, content, construct), and responsiveness of the index, we examined several of its basic properties including the effects of (1) Likert (LK) versus visual analog scale (VAS) formatting[2,3]; (2) blind versus informed presentation[4]; (3) time frame dependency of scores[5]; (4) parametric versus nonparametric forms of analysis[2,3]; (5) signal versus aggregate methods of measurement[6,7]; and (6) weighting and aggregation issues[8]. Originally developed in English, the index has now been translated into several alternative languages for use in more than 25 countries. Recently the WOMAC Index has been recommended as a suitable clinical measure for assessing outcomes in Phase III clinical trials in hip and knee osteoarthritis (OA)[9]. Given its general usage, particularly in multicenter clinical trials, we evaluated a computerized version of WOMAC. The purpose of the study was 2-fold: (1) to validate the computerized format (WOMAC-C) against the original paper format (WOMAC-P), and (2) to assess the feasibility of patients completing WOMAC-C and subsequently transmitting scores by modem from London to a central computer in Toronto.

From the Department of Medicine, The University of Western Ontario, London; Eye Research Institute of Canada, Toronto; and the Volume Investigation Laboratory, The Toronto Western Hospital, Toronto, Canada.

Supported by a grant from Hoffmann-La Roche Limited.

N. Bellamy, MD, MSc, FRCP(Glas, Edin), FRCPC, FACP, Professor of Medicine, Epidemiology and Biostatistics; J. Campbell, BA, Research Assistant, Department of Medicine, University of Western Ontario; J. Stevens, MD, Director, Eye Research Institute of Canada; L. Pilch, Research Assistant; C. Stewart, Research Assistant; Z. Mahmood, Research Assistant, Volume Investigation Laboratory, Toronto Western Hospital.

Address reprint requests to Dr. N. Bellamy, Suite 303, Colborne Building 3, London Health Sciences Centre–Victoria Campus, 375 South Street, London, Ontario, Canada N6A 4G5.

Submitted June 16, 1994 revision accepted May 23, 1997.

MATERIALS AND METHODS

Thirty consecutive consenting outpatients considered to have primary OA of the knee were enrolled in the study. Inclusion criteria included: symptomatic primary OA of at least one knee for at least 3 months; age 45–80 years; American Rheumatism Association (ARA) functional class I–III[10]; radiographic evidence of narrowing of joint space, sclerosis, marginal lipping, bone cysts or osteophyte formation, with a minimum Grade 2 and maximum Grade 3 severity[11]. Exclusion criteria included: inability to comprehend English, ARA functional class IV, prior joint replacement surgery on the study knee. The following disease and demographic variables were recorded: age, sex, disease duration, ARA functional class, radiographic class assessed by comparison against radiographs in the *Atlas of Standard Radiographs*[11]. Patients completed the 2 versions of WOMAC in random order. There was an interval of about 10 min between completion of the 2 versions. Patients did not have access to their prior scores when completing the 2nd version, and since the WOMAC contains 24 questions (answers being made on VAS), it was not considered likely that they could remember their initial scores.

For this study the paper version (WOMAC-P) was the VAS version of WOMAC (syn: WOMAC VA3.0). Patients self-assessed their pain, stiffness, and physical function over the preceding 48 h with respect to their study knee. WOMAC-C was developed as part of a computerized measurement process for longterm studies of the effects of antirheumatic drugs on cartilage. The software was developed by L. Pilch and ran on a NeXT computer. The WOMAC questionnaire was displayed on the color monitor

in a format very similar to that of the original paper version. Patients were instructed to indicate their scores directly on the screen for each of the 24 WOMAC items using a mouse to move the cursor. An assistant (JC) was present to demonstrate how to use the mouse. The VAS were horizontally oriented with the same terminal descriptor format as the paper version. The cursor was in the form of a short box with a central vertical indicator marker. The marker could be set anywhere between the terminal descriptors by mouse manipulation. Patients were given the opportunity to review their cursor placements and revise them if they considered it necessary. The corresponding numerical values between 0 and 100 were automatically generated for each VAS from the exact cursor placements but were not revealed to the patient. Instead, they were stored in London and downloaded each night to the central computer in Toronto.

Statistical analysis was performed using SPSS/PC[12]. Descriptive analyses were conducted on the demographic variables. Mean aggregated WOMAC-C and WOMAC-P subscale scores were compared by paired t tests. Intraclass correlation coefficients (ICC) were calculated between the paper and the computerized aggregated WOMAC subscale scores[13].

RESULTS

Thirty patients completed both versions of the questionnaire. There were 19 women and 11 men with a mean age of 64.8 years (range 46–77). Mean disease duration was 10.8 years (range 4 mo–25 yrs). ARA functional class was graded as Class 2 in 25 patients and Class 3 in 5 patients. The mean and standard deviation (SD) of WOMAC scores for the (P) and (C) formats as well as the mean difference and SD of the difference between the 2 formats are shown in Table 1. Statistically significant differences were observed between aggregate pain ($p = 0.04$) and physical function ($p = < 0.001$) subscale scores using the 2 versions of the questionnaire, but not on the stiffness subscale ($p = 0.44$). The pattern of response even for a single individual was quite variable, responses on WOMAC-C being higher than WOMAC-P on some items but lower on others. In all, there were 720 comparisons (i.e., 24 items and 30 patients). In 56% of instances WOMAC-P scores were greater than WOMAC-C. In 39%, WOMAC-P scores were less than WOMAC-C scores, and in 5% they were identical. In comparisons where a difference was detected, the magnitude of the mean difference was greater when WOMAC-P was higher than when WOMAC-C was higher (Pain 16.3 vs 8.9; Stiffness 13.6 vs 11.5; Function 12.2 vs 8.8; Overall 12.8 vs 9.4). The percentage differences between WOMAC-C and WOMAC-P mean scores were as follows: Pain 8.1%, Stiffness 3.5%, and Physical Function 8.9%. When considered with respect to the corresponding scale lengths (Pain 500 mm, Stiffness 200 mm, Physical Function 1700 mm), the percentage differences in mean scores, i.e., [(WOMAC-P – WOMAC-C/scale length) × 100], were as follows: Pain 3.5%, Stiffness 1.7%, Physical Function 4.1%. The frequency of zero scores was extremely low and similar in the 2 forms of the index (WOMAC-C: Pain 0.01%; Stiffness 0%; Function 0.02%; WOMAC-P: Pain 0%; Stiffness 0%; Function 0.02%. Criterion validity (assessed using ICC), based on aggregate subscale scores between the 2 forms, was excellent (Table 1). No order effects were observed.

Following some simple instructions, no patients had any significant difficulty completing the task, although the majority had no prior experience working with computers. All completed the task in 10–15 minutes. Automatic scoring on site, storage, and remote transmission of data to Toronto were successfully achieved for all patients.

DISCUSSION

The use of valid, reliable, and responsive measures is quintessential to the efficient and successful completion of clinical trials. The WOMAC Osteoarthritis Index, like several other musculoskeletal indices[14], has been rigorously validated in several settings. Its use by several groups of investigators and its incorporation in recent Osteoarthritis Research Society guidelines are indicative of its widespread use in multicenter trials. In such trials the ability to transmit raw data to a central or mainframe computer is of considerable importance. If, in addition, the time consuming task of gathering data from case report forms and/or measuring VAS scores from raw data sheets can be automated, there is further efficiency. Three questions arise: (1) Is the method valid?; (2) What is the feasibility of this approach?; (3) Can the data be irretrievably lost? With respect to this study, we observed excellent correlation, assessed by ICC, between

Table 1. Descriptive statistics for WOMAC-P and WOMAC-C based on aggregated subscale scores.

WOMAC Subscale	Mean	SD	Mean Difference	SD of Difference	t test p value	ICC
Pain						
WOMAC-P	216.2	101.2				
WOMAC-C	198.7	96.0	17.5	44.4	0.04	0.89
Stiffness						
WOMAC-P	95.2	46.8				
WOMAC-C	91.9	41.8	3.3	22.8	0.44	0.87
Physical Function						
WOMAC-P	777.5	357.4				
WOMAC-C	708.6	329.9	68.9	91.0	< 0.001	0.95

SD: standard deviation;
ICC: intraclass correlation coefficient.

WOMAC-C and WOMAC-P scores, suggesting that WOMAC-C is valid and can be used as an alternative to WOMAC-P for clinical trial purposes. We have, in addition, examined agreement, both descriptively and comparatively. The differences in mean scores on the pain and physical function subscales observed between the 2 formats are numerically small and clinically unimportant (i.e., < 10%). However, the slight differences in pain and physical function scores indicate that the 2 forms should not be interchanged during the course of a single trial. The observed difference between WOMAC-P and WOMAC-C scores is accounted for by the fact that WOMAC-C scores were lower (vs WOMAC-P scores) in a greater percentage of responses and also the magnitude of those differences was greater. The reason for computer scores being slightly lower is speculative. It may relate to the near vertical orientation of the computer screen vs the horizontal placement of the paper questionnaire. The use of a mouse rather than a pen may alter the patient's perception of the VAS. Finally, the cursor is broader than the pen marker. Possibly patients line up on the right hand end of the cursor rather than centering it on the intended placement. We evaluated the possibility of a floor effect by comparing the percentage of patients with zero scores. However, these are rare and did not differ between the 2 forms of the WOMAC. Each of these possibilities, while plausible, does not explain how all individuals answered higher on some questions but lower on others (vs WOMAC-P). The feasibility of having patients of varying age, sex, computer literacy, and background complete WOMAC-C has been confirmed in this study. However, we have not studied the extremes of age or the socioeconomically disadvantaged or those less conversant in the English language. Nevertheless, the questionnaire was completed successfully by all patients in a short time (10–15 min). Finally, we were able to automatically score the exact cursor position on the VAS, generate an in-house record, and then automatically download the data to the central computer in Toronto without loss of information. Since a permanent record (hard copy) can also be produced on site before the patient's departure from clinic, the possibility of total data loss can be avoided.

In conclusion, we compared a computerized version of WOMAC against the original paper questionnaire, and observed this to be a valid, feasible, and efficient method of collecting, recording, and transmitting data. Although we have not yet studied the relative sensitivity to change (i.e., responsiveness) of WOMAC-C (vs WOMAC-P), this study has implications for the conduct of future clinical trials in OA, particularly those using multiple centers. We are currently completing conversion of the software to run on IBM compatible computers. This adaptation should provide broader access to researchers working in IBM formats.

ACKNOWLEDGMENT

The authors thank Larry Stitt, MSc, for statistical support and Hoffmann-La Roche Limited (Canada) for supporting the development of the original software and the recent conversion to IBM PC.

REFERENCES

1. Bellamy N: Osteoarthritis — An evaluative index for clinical trials (MSc thesis). Hamilton, Canada: McMaster University, 1982.
2. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt L: Validation study of WOMAC: A health status measure for measuring clinically important patient relevant outcomes following total hip or knee arthroplasty in osteoarthritis. *J Orthop Rheumatol 1988;*1:95–108.
3. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt L: Validation study of WOMAC: A health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol 1988;*15:1833–40.
4. Bellamy N, Goldsmith CH, Buchanan WW, Campbell J, Duku E: Prior score availability: Observations using the WOMAC Osteoarthritis Index. *Br J Rheumatol 1991;*30:150–1.
5. Griffiths G, Bellamy N, Kean WF, Campbell J, Gerecz-Simon E: A study of the time frame dependency of the WOMAC Osteoarthritis Index. *Inflammopharmacology 1993;*2:85–7.
6. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Duku E: Signal measurement strategies: Are they feasible and do they offer any advantage in outcome measurement in osteoarthritis? *Arthritis Rheum 1990;*33:729–45.
7. Burr S, Bellamy N, Campbell J, *et al:* A comparative study of signal versus aggregate methods of outcome measurement based on the WOMAC Osteoarthritis Index. *J Rheumatol 1994;*21:2106–12.
8. Bellamy N, Wells GA, Campbell J: Relationship between severity and clinical importance of symptoms in osteoarthritis. *Clin Rheumatol 1991;*10:138–43.
9. Osteoarthritis Research Society Task Force: Design and conduct of clinical trials in patients with osteoarthritis: Recommendations from a task force of the Osteoarthritis Research Society. *Osteoarthritis Cart 1996;*4:217–43.
10. Steinbrocker O, Traeger CH, Batterman RC: Therapeutic criteria in rheumatoid arthritis. *JAMA 1949;*140:659–62.
11. Kellgren JK, Lawrence JS: *The Epidemiology of Chronic Rheumatism-Atlas of Standard Radiographs,* vol 2. Philadelphia: FA Davis, 1963.
12. SPSS/PC Users Guide. Chicago: SPSS Inc., 1984.
13. Fleiss JL: The measurement of interrater agreement. In: *Statistical Methods for Rates and Proportions.* New York: John Wiley & Sons, 1981:212–36.
14. Bellamy N: *Musculoskeletal Clinical Metrology.* Dordrecht: Kluwer Academic, 1993:11–36.

A Comparative Study of Telephone versus Onsite Completion of the WOMAC 3.0 Osteoarthritis Index

NICHOLAS BELLAMY, JANE CAMPBELL, JONATHAN HILL, and PHILIP BAND

ABSTRACT. ***Objective.*** Outcome assessment in clinical trials using the Western Ontario and McMaster University (WOMAC 3.0) Osteoarthritis Index is traditionally achieved through self-administration of the Index. However, in other areas of clinical measurement, telephone administration has been shown to be a reliable method of acquiring data that are both accurate and complete. To address this issue in knee osteoarthritis (OA), we conducted a comparative study of telephone administration by interviewer of WOMAC LK3.0 versus onsite self-completion at the hospital.

Methods. Fifty consenting patients with knee OA were randomized to complete the WOMAC LK3.0 Index by telephone interview one day, followed by onsite completion the following day, or vice versa. Neither patients nor interviewers had access to any prior scores.

Results. The mean age of the 50 patients was 66.3 years (range 44–82); 34 (68%) were female and 16 (32%) male. There was excellent agreement between the mean office and telephone scores, with mean differences for the WOMAC LK3.0 pain, stiffness, and function subscale scores and total score of 0.09, 0.12, 0.78, and 0.98, respectively. These differences were well within the respective protocol defined equivalence criteria of ± 1.7, ± 0.9, ± 6.4, and ± 9.1, and represented differences from office scores of 0.9, 2.6, 2.4, and 2.2%, respectively.

Conclusion. The use of telephone interviews for the WOMAC LK3.0 Index is a valid method of obtaining OA outcome measurements. These observations have important implications for designing data acquisition strategies for future OA clinical trials and for longterm observational studies. (J Rheumatol 2002;29:783–6)

Key Indexing Terms:
OUTCOME ASSESSMENT OSTEOARTHRITIS
WOMAC SURVEY TELEPHONE

The use of telephone contact to collect information on patient outcomes is well founded in the literature. This method has been used in several clinical studies and has been found to be a valid tool for patient outcome assessment[1-5]. However, searching under the MeSH headings "telephone," "survey," "VAS," "WOMAC," "osteoarthritis," and "outcomes," no published data could be found specifically validating the use of data collected by telephone in osteoarthritis (OA) outcome studies. The Western Ontario and McMaster University (WOMAC) Osteoarthritis Index has been in use for 19 years[6-8], but to date there has been no formal assessment of the Index comparing telephone administration by interviewer against the usual approach of self-completion onsite in the clinic or doctor's office. Verification of the validity of telephone data, in comparison to visit-collected data, would be useful for designing data acquisition strategies for future OA clinical trials, as well as for longterm longitudinal studies. This study was undertaken to validate telephone assessment of the WOMAC LK3.0 Index by comparing the results of telephone interviews versus onsite assessments.

From the Centre of National Research on Disability and Rehabilitation Medicine (CONROD), Brisbane, Australia; the University of Western Ontario, London, Ontario, Canada; and Matrix Biology Institute, Ridgefield, New Jersey, USA.

Sponsored by Biomatrix, Inc.

N. Bellamy, MD, MSc, FRCPC, FRCP (Glas and Edin), FACP, Professor, Director, Centre of National Research on Disability and Rehabilitation Medicine (CONROD), University of Queensland; J. Campbell, BA, CCRC, Department of Surgery, University of Western Ontario; J. Hill, BSc, Department of Microbiology and Immunology, University of Western Ontario; P. Band, PhD, Matrix Biology Institute.

Address reprint requests to Prof. N. Bellamy, Centre of National Research on Disability and Rehabilitation Medicine, University of Queensland, Department of Medicine, Level 3, Mayne Medical School, Herston Road, Brisbane, Queensland 4006, Australia.

Submitted February 15, 2001; revision accepted October 17, 2001.

MATERIALS AND METHODS

This was a single center, outpatient study, designed to compare patient responses to the WOMAC LK3.0 instrument administered by telephone and during an office visit. It was conducted as part of a larger study, comparing telephone vs office administration of 3 visual analog measures of pain and function (TPQ-VAS) and a 5 point Likert scaled patient global assessment (PGA) question whose performance we are evaluating. In each case, the WOMAC LK3.0 was the first questionnaire completed, since we wished to avoid reactivity on WOMAC LK3.0 scores. The WOMAC LK3.0 is a tridimensional joint targeted, patient centered questionnaire containing 5 pain, 2 stiffness, and 17 physical function questions, responses to which are scaled on 5 point (none, mild, moderate, severe, extreme) adjectival scales resulting in subscale score sizes of 0–20 for pain, 0–8 for stiffness, 0–68 for physical function, and 0–96 for total score. Male and female outpatients from the investigator's practice were initially contacted by one of the following methods: (1) by letter requesting their participation in the study; (2) by telephone call; or (3) during their visit to an outpatient clinic. Screening procedures included a review of the inclusion criteria. Patients meeting the following criteria were eligible for entry into the

study: over 35 years of age, inclusive; radiographic evidence of knee OA[9]; ability to understand English; and willingness to participate in the study according to the protocol. Because of the noninvasive nature of this study, there were no explicit exclusion criteria. Patients who wished to enroll received a blank copy of the WOMAC LK3.0 questionnaire for reference during the telephone interview. The first 50 patients from the investigator's patient population who consented to participate were selected for study. After enrolling in the study, patients were randomized to one of 2 groups. Group 1 completed questionnaires by telephone on Day 1, followed on Day 2 by questionnaire completion in the office. Group 2 completed questionnaires in the office on Day 1, followed by telephone completion on Day 2. All interviews, whether by telephone or in the office, were supervised by an evaluator trained in the administration and evaluation of the WOMAC LK3.0 Index. At the beginning of the study, patients were asked to select the knee with the worst OA symptoms. This was then designated their study knee. The WOMAC LK3.0 questionnaire contained written instructions for completion of the form. Patients were instructed to rate the severity of the pain, stiffness, and physical disability they had experienced in their study knee over the previous 4 weeks. The WOMAC LK3.0 was self-administered in the office. During the telephone assessment, the evaluator read each WOMAC LK3.0 question, and then recorded the patient's response on the blank questionnaire. Neither the patient nor the evaluator had access to any prior WOMAC LK3.0 scores.

Patients were free to withdraw from the study at any time. However, because of the short duration and noninvasive nature of the trial, no explicit provisions for removal of patients from assessment were included in the study protocol. This study did not involve drug administration or other invasive procedures, and patients were not requested to make any changes to their existing treatment regimen prior to or during the assessment period. Prior and concomitant therapies were not expected to have a material effect on the results because of the short retest interval, and therefore concomitant therapies were neither recorded nor controlled.

Statistical issues. As noted, this study was conducted within the framework of a larger study. As a consequence, sample size calculations were performed to estimate the minimum sample size required to establish with 95% confidence that TPQ-VAS telephone values did not differ by more than 15 mm from office administered values. Statistical analyses for equivalence were performed using analysis of variance (ANOVA) (alpha = 0.05) to establish the upper and lower limits of the relevant confidence interval[10]. Twenty percent has been used in the arthritis literature as a threshold value for declaring a clinically important difference[11], and earlier studies have established standard deviations for VAS measures of pain of about 15 mm[12,13]. Equivalence was to be inferred if the 95% confidence limits for the differences between the office and telephone scores were within ± 20% of the mean office scores. Based on these assumptions, a sample size of 50 patients was determined to be adequate for the required statistical power. This sample size was also considered adequate for evaluating similar issues for the WOMAC LK3.0 and the PGA (i.e., equivalence was to be inferred if the 95% confidence limits for the differences between the office and telephone scores were within ± 20% of the mean office scores).

After completion of the questionnaire stage of the study, data were entered from the patient questionnaires into a database and subjected to quality assurance procedures that were double verified, providing 100% verification for the key OA outcome measurements. There was no a priori reason to suspect that performing the assessment in the office first would influence the outcome of the subsequent telephone interview and vice versa. However, because of this, it was necessary to ensure that results from the office visit and the telephone interview were consistent, regardless of order of presentation.

For each outcome measure, ANOVA statistics were computed for the telephone and office scores, classifying them with respect to those obtained first and those obtained second. The first and second office scores and the first and second telephone scores were to be pooled if p values for a given outcome measure were greater than 0.05. Comparisons were made using the pooled error estimate obtained from the ANOVA. Scores for outcome measures that satisfied the pooling criterion were pooled and evaluated for equivalence as follows. WOMAC LK3.0 data were treated as continuous normally distributed data for the purposes of the analyses. Difference scores for each WOMAC LK3.0 component were calculated on a by-patient basis by subtracting the value determined on the telephone from the corresponding office visit value. These differences were then subjected to statistical analysis based on the principles advanced by Bland and Altman[10], and using the SAS program PROC Means.

RESULTS

An initial cohort of 50 patients was selected for the study. One patient, who did not have a knee radiograph on file to provide radiographic evidence of OA, was subsequently dropped from the study. One other patient, randomized to telephone assessment on Day 1, was assessed on Day 2 at her home instead of in the office and was also dropped from the study. Data from these patients were not included in the database. Two additional patients were recruited to achieve the desired study cohort of 50 patients. Of the final 50 patients, 22 were randomized to Group 1 (telephone first) and 28 were randomized to Group 2 (office first).

In violation of the protocol, 5 patients did not complete questionnaires on consecutive days. Four of these patients completed the second assessment on Day 3, rather than Day 2. One patient completed the second assessment on Day 5 instead of Day 2. However, because patients were rating the condition of their study knee over a 4 week timeframe, these deviations were not considered to be serious violations. Further, any effect of these more extensive intervals would be expected to decrease rather than increase observed levels of agreement. Therefore, the scores for these patients were included in the analysis.

The mean age of the 50 randomized patients was 66.3 years (range 44 to 82). Of these 50 patients, 34 (68%) were female and 16 (32%) male. The severity of the patients' OA ranged from Grade I to IV, as graded by Kellgren-Lawrence radiographs[9]: Grade I 20%, Grade II 38%, Grade III 40%, Grade IV 2%.

The significance of the effect of order (first or second) on assessment method (office or telephone interview) was analyzed by ANOVA. The results are provided in Table 1. The results indicate there were no statistically significant differences attributed to the order of questionnaire administration for any of the outcome measures. This means that office and telephone scores from those patients randomized to the telephone-first questionnaire were statistically equivalent to the corresponding office and telephone scores for patients in the office-first group. Therefore, the results from the 2 groups of patients were combined for the analysis of success criteria.

As described above, the WOMAC LK3.0 telephone and office visit methods were considered equivalent if the combined 2 sided 95% CI for the differences between methods were within ± 20% of the mean office scores. Although all patients were symptomatic, this study included patients with a wide range of OA symptoms. As a result,

Table 1. Analysis of sequence effects.

Outcome Measure	Mean Office Scores (SD)			Mean Telephone Scores (SD)		
	First, n = 28	Second, n = 22	p	First, n = 22	Second, n = 28	p
WOMAC Pain	9.11 (3.17)	8.16 (3.15)	0.424	8.15 (3.19)	8.96 (2.77)	0.491
WOMAC Stiffness	5.00 (1.70)	3.95 (1.68)	0.120	3.91 (1.72)	4.82 (1.72)	0.174
WOMAC Function	34.71 (11.46)	28.96 (12.37)	0.212	27.65 (11.53)	34.36 (10.52)	0.147
WOMAC Total	48.82 (14.70)	41.07 (16.41)	0.205	39.71 (15.52)	48.14 (14.05)	0.168

zero scores on some, but never on all outcome measures, did occur for the outcome measures of specific patients. The primary success criteria evaluated in this study were analyzed by subtracting the telephone score from the corresponding office visit score for each WOMAC component. The differences were then subjected to a statistical analysis using the SAS program PROC Means. The results of this analysis are provided in Table 2. There was excellent agreement between the mean office and telephone scores, with mean differences for the WOMAC LK3.0 pain, stiffness, function, and total scores of 0.09, 0.12, 0.78, and 0.98, respectively. These differences are also well within the protocol-defined equivalence criteria of ± 1.7, ± 0.9, ± 6.4, and ± 9.1, respectively, for pain, stiffness, physical function, and total WOMAC LK3.0 scores, and represent differences from office scores of 0.9, 2.6, 2.4, and 2.2%, respectively.

The exact time taken to complete the questionnaires was not measured. However, completion times estimated by evaluators were between 5 and 10 minutes.

DISCUSSION

Telephone contact to collect information on patient outcomes has been used extensively in clinical studies and has been found to be a valid tool for patient outcome assessment. The ability to use telephone contacts to obtain OA patient health assessments would be of great benefit to those patients for whom office visits can be both difficult and inconvenient.

This study was designed to validate the use of WOMAC LK3.0 telephone interviews as a measure of OA study outcome by comparing the results of telephone versus office visit assessments. The WOMAC LK3.0 demonstrated excellent agreement between the mean office and telephone scores, with mean differences for the WOMAC LK3.0 outcome measures ranging from 0.09 to 0.98. The 95% CI were well within established equivalence criteria of ± 20% of the mean office scores for the respective outcome measures. The telephone administered WOMAC can therefore be considered as validated by the criteria established in the protocol. This demonstration of clinical and statistical equivalence provides the first evidence that the use of telephone interviews using the WOMAC LK3.0 Index is a valid method of obtaining OA outcome measurements.

In recognizing the following potential limitations, it is acknowledged that the results of this study are directly generalizable only to individuals and groups having similar characteristics to this group of patients. We have not specifically addressed issues peculiar to the elderly, those not fluent in English, institutionalized patients, those not under medical care, or those of low education or socioeconomic status. Furthermore, the study population was not constituted to permit such an analysis. It is notable that Bombardier, *et al* have successfully administered the WOMAC Index to OA patients seen by family physicians, supporting the contention that this method of administration is feasible in a community based setting[14]. With respect to memory effects, we have experience with varying the timeframe of the WOMAC Index[15] and the interval between pain assessments[16,17], as well as performing repeated assessments

Table 2. WOMAC LK3.0 outcome measures: analysis of differences between telephone and office assessments (n = 50).

	WOMAC Pain	WOMAC Stiffness	WOMAC Function	WOMAC Total
Mean office scores (SD)	8.69 (3.16)	4.54 (1.75)	32.18 (12.09)	45.41 (15.80)
Mean telephone scores (SD)	8.61 (2.96)	4.42 (1.76)	31.41 (11.47)	44.43 (15.16)
Mean difference	0.09	0.12	0.78	0.98
SD, difference	1.83	0.92	3.43	4.05
Paired t, difference	0.33	0.92	1.60	1.71
Prob > T, difference	0.74	0.36	0.12	0.09
Lower 95%, difference	−0.44	−0.14	−0.20	−0.17
Upper 95%, difference	0.61	0.38	1.75	2.13
Protocol defined equivalence criteria ± 20% of mean office score	±1.74	±0.91	±6.44	±9.08

of health status on the same patients without time/patient interactions[18]. While the potential for memory effects deserves recognition, our experience in dissecting circadian rhythmicity in knee OA[18], hand OA[19], and rheumatoid arthritis[20] strongly suggests that patients can detect even small changes in symptom intensity, and memory effects are negligible or absent.

We did not collect quantitative data on patient and interviewer experience evaluations, or the costs and time of administration. The costs of conducting telephone interviews are mostly attributable to the costs of recruiting, training, and retaining interviewers and charges for telephone service and line usage. The former does not impose a high skill requirement, while the latter offers opportunities to negotiate favorable rates if long distance or high volume usage is contemplated. Offsetting this is the greater convenience and cost savings to the patient of not having to leave home, and the opportunity to collect community based rather than specialist based information.

It should be noted that the administration mode-dependent differences detected are very small and in magnitude fall below the values of published definitions of minimum perceptible clinical improvement[21], minimum clinically important difference[22], and responder criteria for OA clinical trials[23]. This suggests that the WOMAC Index is capable of detecting meaningful alterations in health status, when administered by telephone. It is worth reiterating that in this study patients were provided with a blank copy of the WOMAC Index for reference during the telephone interview.

These issues notwithstanding, we believe our findings can be applied to designing data acquisition strategies for future OA clinical trials and longterm observational studies. From a research and regulatory perspective, it facilitates the completeness and speed of data acquisition and transfer, and from the patient's perspective allows an accurate assessment of OA status without the inconvenience and physical demands associated with office visits.

REFERENCES

1. Smith PM, Illig SB, Fiedler RC, Hamilton BB, Ottenbacher KJ. Intermodal agreement of follow-up telephone function assessment using the Functional Independence Measure in patients with stroke. Arch Phys Med Rehabil 1996;77:431-5.
2. Korner-Bitensky N, Wood-Dauphine S. Barthel Index information elicited over the telephone. Is it reliable? Am J Phys Med Rehabil 1995;74:9-18.
3. Rohde P, Lewinsohn PM, Seeley JR. Comparability of telephone and face-to-face interviews in assessing axis I and II disorders. Am J Psychiatry 1997;154:1593-8.
4. Balas EA, Jaffrey F, Kuperman GJ, et al. Electronic communication with patients. Evaluation of distance medicine technology. JAMA 1997;278:152-9.
5. McGrory BJ, Shinar AA, Freiberg AA, Harris WH. Enhancement of the value of hip questionnaires by telephone follow-up evaluation. J Arthroplasty 1997;12:340-3.
6. Bellamy N. Osteoarthritis — An evaluative index for clinical trials [MSc thesis]. Hamilton: McMaster University; 1982.
7. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988;15:1833-40.
8. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes following total hip or knee arthroplasty in osteoarthritis. J Orthop Rheumatol 1988;1:95-108.
9. Kellgren J, Lawrence J. Atlas of standard radiographs: The epidemiology of chronic rheumatism. Vol. 2. Oxford: Blackwell Scientific; 1963.
10. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986; 8:307-10.
11. Bellamy N. Standardised procedures for outcome measurement and parameters for calculating sample size for antirheumatic drug studies. Musculoskeletal clinical metrology. Dordrecht: Kluwer Academic Publishers; 1993:214-24.
12. Bellamy N, Carette S, Ford PM, et al. Osteoarthritis antirheumatic drug trials. III. Setting the delta for clinical trials — results of a consensus development (Delphi) exercise. J Rheumatol 1992;19:451-7.
13. Felson D, Anderson JJ, Boers M, et al. American College of Rheumatology preliminary definition of improvement in rheumatoid arthritis. Arthritis Rheum 1995;38:727-35.
14. Bombardier C, Maguire L, Li L, Maetzel A, Jansz G, The CHAP Team. Characteristics of patients with osteoarthritis seen by family physicians: results of a telephone survey in Ontario [abstract]. J Rheumatol 2001;28:1416.
15. Griffiths G, Bellamy N, Kean WF, Campbell J, Gerecz-Simon E. A study of the time frame dependency of responses to the WOMAC Osteoarthritis Index. Inflammopharmacology 1993;2:85-7.
16. Bellamy N, Goldsmith GH, Buchanan WW, Campbell J, Duku E. Prior score availability: Observations using the WOMAC Osteoarthritis Index. Br J Rheumatol 1991;30:150-1.
17. Bellamy N, Carette S, Ford PM, et al. Osteoarthritis antirheumatic drug trials. I. Effects of standardization procedures on observer dependent outcome measures. J Rheumatol 1992;19:436-43.
18. Bellamy N, Sothern RB, Campbell J. Rhythmic variations in pain perception in osteoarthritis of the knee. J Rheumatol 1990; 17:364-72.
19. Bellamy N, Campbell J, Sothern R, Soucy E, Flynn J, Buchanan WW. Circadian rhythm in pain, stiffness and manual dexterity in osteoarthritis of the hand: relationship between discomfort and disability. International Congress on Chronobiology, Paris, September 7-11, 1997. Paris Chronobiology International 1997;24 Suppl 1:16.
20. Bellamy N, Sothern RB, Campbell J, Buchanan WW. Circadian rhythm in pain, stiffness and manual dexterity in rheumatoid arthritis: relation between discomfort and disability. Ann Rheum Dis 1991;50:243-8.
21. Ehrich EW, Davies GM, Watson DJ, Bolognese JA, Seidenberg BC, Bellamy N. Minimal perceptible clinical improvement with the Western Ontario and McMaster Universities Osteoarthritis Index questionnaire and global assessments in patients with osteoarthritis. J Rheumatol 2000;27:2635-41.
22. Bellamy N, Carr A, Dougados M, Shea B, Wells G. Towards a definition of "difference" in osteoarthritis. J Rheumatol 2001;28:427-30.
23. Dougados M, LeClaire P, van der Heijde D, Bloch DA, Bellamy N, Altman RD. Special article: Response criteria for clinical trials on osteoarthritis of the knee and hip: A report of the Osteoarthritis Research Society International Standing Committee for Clinical Trials Response Criteria Initiative. Osteoarthritis Cartilage 2000;8:395-403.

Osteoarthritis and Cartilage (2002) 10, 479–481

1063–4584/02/$35.00/0
doi:10.1053/joca.2002.0807, available online at http://www.idealibrary.com on IDEAL®

Osteoarthritis and Cartilage
OsteoArthritis Research Society International

Clinical evaluation of the WOMAC 3.0 OA Index in numeric rating scale format using a computerized touch screen version

R. Theiler*, J. Spielberger*, H. A. Bischoff†, N. Bellamy‡, J. Huber* and S. Kroesen*
**Division of Rheumatology, Cantonal Hospital Aarau, Switzerland; †Division of Rheumatology, Immunology and Allergy, the Robert B. Brigham Arthritis and Musculoskeletal Clinical Research Centre, Brigham and Women's Hospital, Boston, U.S.A.; ‡Centre of National Research on Disability and Rehabilitation Medicine (CONROD), Department of Medicine C Floor, Clinical Sciences Building, Royal Brisbane Hospital, Brisbane, Australia*

Summary

Background: The Western Ontario and McMaster Universities (WOMAC) Osteoarthritis Index is a previously described self-administered questionnaire covering three domains: pain, stiffness and function. It has been validated in patients with osteoarthritis (OA) of the hip or knee in a paper-based format.

Aim: To validate the WOMAC 3.0 using a numerical rating scale in a computerized touch screen format allowing immediate evaluation of the questionnaire. In the computed version cartoons, written and audio instruments were included in order facilitate application.

Methods: Fifty patients, demographically balanced, with radiographically proven primary hip or knee OA completed the classical paper and the new computerized WOMAC version. Subjects were randomized either to paper format or computerized format first to balance possible order effects.

Results: The intra-class correlation coefficients for pain, stiffness and function values were 0.915, 0.745 and 0.940, respectively. The Spearman correlation coefficients for pain, stiffness and function were 0.88, 0.77 and 0.87, respectively.

Conclusion: These data indicate that the computerized WOMAC OA index 3.0 is comparable to the paper WOMAC in all three dimensions. The computerized version would allow physicians to get an immediate result and if present a direct comparison with a previous exam.

Key words: Patient self-assessment, Electronic WOMAC 3.0, Electronic data capturing (EDC), QUALITOUCH method.

Introduction

The Western Ontario and McMaster Universities (WOMAC) Osteoarthritis Index was developed for standardized assessment of osteoarthritis (OA) symptoms in hip and/or knee joints[1]. It is composed of 24 questions covering three dimensions: pain (five questions), stiffness (two questions) and function (17 questions). The WOMAC OA Index has been extensively tested for validity, reliability, feasibility and responsiveness for measuring changes after different OA interventions[2,3].

Consensus was reached at the third conference on outcome measures in rheumatoid arthritis clinical trials (OMERACT III). An OA research society (OARSI) task force dealing with outcome measurement in OA clinic trials decided that the WOMAC OA Index as a disease specific questionnaire is recommended for core set assessment in OA clinical trials for knee and hip OA[4].

In the present study we used a simplified computer touch screen format that could be applied in senior citizens or non-computer skilled individuals by offering a multimedia presentation: cartoon, written and spoken (Fig. 1). We hypothesized that this method is a reliable way for assessing the WOMAC if compared with the original paper version. The scaling of the computerized WOMAC in a previous study[5] was a visual analog scale. In contrast we used the same questionnaire but answered a numeric rating scale format.

The development of a computer version of the WOMAC index application was of research and clinical interest. This application could improve the quality of data collection in clinical trials by computer-based direct data harvesting. In addition it could simplify its use both in the research setting and in daily clinical practice.

Materials and methods

Fifty consecutive seen outpatients with radiographically proven primary OA were invited to complete both a paper format and a computerized touch screen format of the WOMAC OA Index. The demographics of these patients can be seen in Table I. The following inclusion criteria were employed: symptomatic OA at least in one joint of the lower limb with symptoms lasting for at least 3 months and ability to comprehend the German language. Exclusion criterion was prior joint replacement on the study joint.

Received 23 August 2001; accepted 24 January 2002.
Address correspondence to: Theiler R. M.D., Rheumatology Unit, Cantonal Hospital Aarau, Buchserstrasse 1, CH 5000 Aarau, Switzerland.
This work was supported by: Hugo and Elsa Isler-Stiftung.

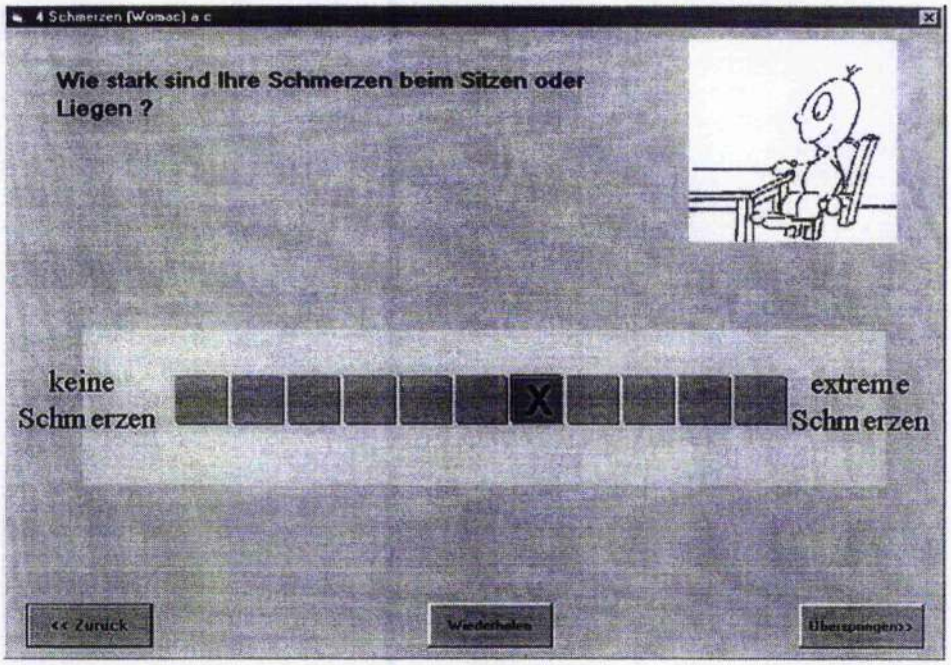


Fig. 1. The screen display of the question 4. Translation: Wie stark sind Ihre Schmerzen beim Sitzen oder Liegen?=How much pain do you have in sitting or lying. Keine Schmerzen=no pain. Extreme Schmerzen=extreme pain. Zurück=back. Wiederholen=repeat question. Überspringen-next.

Patients either completed the paper form first or the computerized version first. The mean time interval between completing the computerized version and the paper format of the WOMAC OA Index was 16 min. Patients were not able to see their prior scores. The German paper version of WOMAC 3.0 OA Index using a numeric rating scale format, which has been validated, was used[6]. For the computerized version audio and visual cues were presented on a 34.3 cm diameter screen. The questions were answered by touching one of the squares of the numeric rating scale on the computer screen. By using five buttons on screen the patient could get help and was able to move one question forward or backward. It was not possible to leave one question unanswered. Furthermore the help function self activated after 15 s inactivity and presented the next possible steps to the patient. The software was developed by a private programming company, as was the purpose built touch screen computer. This data capturing method was called the QUALITOUCH method.

Table I
Demographie: validation WOMAC NR 3.0

Male	29 patients
Female	21 patients
Age (range)	61 (34–79) years
Knee right	12 joints
Knee left	10 joints
Hip right	8 joints
Hip left	9 joints
Knee both side	5 joints
Hip both sides	5 joints
Mean time difference between completing the forms	15 min

The numeric rating scale scores 0 (best) and 10 (worst) health. Therefore,the maximal aggregated score for pain, stiffness and function was 50, 20 and 170, respectively.

A block randomization with a block size of four was used for creating the two groups. Descriptive statistics included the mean of the aggregated scores, the standard deviation and the mean difference between the scores of the paper and computerized version. Agreement was assessed using intraclass correlation of Spearman's rank correlation coefficient.

Results

Fifty patients completed both versions of the questionnaire. There were 21 female and 29 male with a mean age of 50.5 years (range 34–79). Age and gender was balanced between the groups. The means and standard deviations (S.D.) of WOMAC scores for the (P) and (C) formats as well as the mean difference and S.D. of the difference between the two formats are shown (Table II). No statistically significant differences were observed for aggregate pain (P=0.9), stiffness (P=0.6) and physical function (P=0.9) subscale sores using the two versions of the questionnaire.

The relative (%) differences in mean scores based on scale length ((WOMAC-C−WOMAC-P/scale length]×100), were 0.07 for pain, 2.18 for stiffness and 0.42 for physical function. A tendency towards zero scores was observed in the three different sections of the WOMAC index. Criterion validity, assessed and based on aggregate subscale scores between the two formats, was excellent (Table II). No order effects were observed.

Following standardized instructions, no patient had significant difficulty completing the task, although majority had

Table II
Descriptive statistics for WOMAC-P and WOMAC-C based on aggregated subscale scores

WOMAC subscale	Mean score	s.d.	Mean score difference	s.d. of difference	*t*-test *P* value	ICC
Pain						
WOMAC-P	15.64	10.3	0.04	4.28	0.98	0.91
WOMAC-C	15.68	10.4				
Stiffness						
WOMAC-P	8.28	5.4	0.48	3.76	0.65	0.74
WOMAC-C	7.80	5.0				
Function						
WOMAC-P	53.8	36.8	0.78	12.42	0.91	0.94
WOMAC-C	54.5	35.9				

s.d.: standard deviation.
ICC: Intra class correlation.

no prior experience in working with the computer. All completed the tasks in 10–15 min.

Discussion

The WOMAC questionnaire is a frequently used outcome measure in patients with lower limb OA. To our knowledge this is the first published report about a touch screen version of the WOMAC questionnaire. The present study shows that the touch screen format is reliable if compared with the original paper version.

In this computerized WOMAC format the questions are shown in a cartoon, written and spoken. This QUALITOUCH method could improve and facilitate the patient's understanding. The questions are answered by touching the screen directly. Thereby neither keyboard nor mouse is necessary for working with this computer. This may be important if non-computer skilled persons or senior citizens are using the computerized version of the WOMAC questionnaire.

Future applications are seen in the establishment of regional, national or international databases of OA patients by connecting the computer to an Internet application. In daily clinical practice patients could be asked to do the WOMAC index on the portable computer while waiting for e.g. a physiotherapy assessment. The immediate evaluation with a graphical display could be used as a patient education tool in the process of rehabilitation, especially if it is used in a longitudinal perspective to track the long-term outcome. In addition the computer version might be a tool for future quality management projects e.g. in orthopedic surgery.

In conclusion, the computerized version of the WOMAC seems to be a valid and feasible instrument for outcome measurement in clinical OA trials.

Acknowledgments

Thanks to Dr Margot vonDechend for support in data collection. This work was supported by an unrestricted research grant of the Hugo and Elsa Isler Foundation, Aarau. Correspondence to: Robert Theiler, MD, PD, Depts. of Rheumatology and Rehabilitation, City Hospital Triemli, 8063 Zurich, Switzerland. Email: robert.theiler@bluemail.ch

References

1. Bellamy N, Buchanan WW. A preliminary evaluation of the dimensionality and clinical importance of pain and disability in osteoarthritis of the hip and knee. 1986;5:231–41.
2. Bellamy N, Goldsmith CH, Campbell J. Validation study of WOMAC: A health status instrument for measuring clinically-important patient-relevant outcomes following total hip or knee arthroplasty in osteoarthritis. J Orthop Rheum 1988;1:95–108.
3. Bellamy N, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. Rheumatol 1988;15:1833–40.
4. Bellamy N, Boers M, Brooks P, Strand V, Tugwell P, Altman R, *et al.* Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. J Rheumatol 1997;24:799–802.
5. Bellamy N, Campbell J, Stevens J, Pilch L, Stewart C, Mahmood Z. Validation study of a computerized version of the Western Ontario and McMaster Universities VA3.0 Osteoarthritis Index. J Rheumatol 1997;24:2413–15.
6. Stucki G, Stucki S, Michel BA, Tyndall AG, Dick W, Theiler R. Evaluation of a German version of WOMAC (Western Ontario and McMaster Universities) Arthrosis Index. Z Rheumatol 1996;55:40–9.

OsteoArthritis and Cartilage (2004) 12, 912–916

doi:10.1016/j.joca.2004.08.006

Osteoarthritis and Cartilage

International Cartilage Repair Society

OARSI OSTEOARTHRITIS

Responsiveness of the electronic touch screen WOMAC 3.1 OA Index in a short term clinical trial with rofecoxib[1]

R. Theiler†*, H. A. Bischoff-Ferrari‡, M. Good§ and N. Bellamy‖

† *Rheumatology Clinic and Institute for Physical Medicine and Rehabilitation, Stadtspital Triemli, Zurich, Switzerland*

‡ *Division of Rheumatology, Immunology and Allergy, Robert B. Brigham Arthritis and Musculoskeletal Clinical Research Center, Boston, MA, USA*

§ *Target BioScience Ltd., Rüschlikon-Zurich, Switzerland*

‖ *Centre of National Research on Disability and Rehabilitation Medicine (CONROD), Mayne Medical School, Floor 3, Herston RD, Brisbane, Australia*

Summary

Background: The Western Ontario and McMaster Universities (WOMAC) Osteoarthritis Index is a self-administered validated questionnaire for patients with osteoarthritis (OA) of the hip or knee. The electronic touch screen version of the WOMAC (e-WOMAC) has been previously shown to be highly correlated with the original paper format. However, whether the e-WOMAC would be suitable for monitoring the effects of drug treatment is unknown.

Aim: To validate the longitudinal use of the e-WOMAC questionnaire and its ability to detect changes in WOMAC-scores induced by drug treatment in outpatient care.

Methods: Fifty-three outpatients, men and women (mean age: 64 years; SD ± 9.5), with symptomatic osteoarthritis of hip or knee were included in an open label study with rofecoxib. At three visits over 3 weeks, responsiveness of the WOMAC 3.1 regarding the three subscales, pain, stiffness and function, were compared for the original paper format and the computer touch screen format (QUALITOUCH®) using a Likert scale. WOMAC scores were transformed to the 0–100 scale. ANOVA for repeated measures was used for analysis and effect sizes by subscale were compared for both formats.

Results: Responsiveness for all three subscales was similar between formats. In both formats, pain and stiffness were significantly reduced with rofecoxib as early as 7 days, while functional ability was significantly increased ($P < 0.01$ for all aggregate subscale scores) with continuing improvement until the end of study. The effect sizes by subscale between Day 1 and 21 were not statistically different between the paper and the electronic version of the questionnaire and showed similar clinically meaningful improvements in WOMAC scores over 3 weeks.

Conclusion: In this longitudinal intervention study, the e-WOMAC OA Index 3.1 showed similar responsiveness in detecting clinically meaningful changes than the original paper format.

Key words: Patient self-assessment, Electronic WOMAC 3.1, Electronic data capturing (EDC), QUALITOUCH method, Rofecoxib.

Introduction

The Western Ontario and McMaster Universities (WOMAC) Osteoarthritis Index was developed for standardized assessment of osteoarthritis (OA) symptoms in hip and/or knee joints[1]. The WOMAC OA Index covers the domains of pain, stiffness and function in 5, 2 and 17 questions, respectively. This Index has been extensively validated for measuring changes after different interventions in patients with OA[2] and is the most widely recommended disease-specific questionnaire for core set assessment in clinical trials for knee and hip OA established by the Osteoarthritis Society Task Force, as proposed at the third conference on outcome measures in rheumatoid arthritis clinical trials (OMERACT III)[3].

In daily clinical practice the WOMAC questionnaire is a suitable tool for optimizing patient monitoring as the data are directly provided by the patient and are very reproducible. However, the paper format does not allow for an immediate display of results. The e-WOMAC was designed to improve patient monitoring by its simple design and provides the opportunity to discuss results with the patients or the team that takes care of the patients in a timely fashion, as results are available immediately and can be shared electronically[4]. Another advantage of the e-WOMAC may be its presentation format, where each question is displayed as text and a situational cartoon, and are verbalized over the loudspeaker (QUALITOUCH®

[1]This study was supported by a grant of MSD-Chibret AG, Glattbrugg, Switzerland.

*Author correspondence and reprint requests to: PD Dr. Med. Robert Theiler, Head of Rheumatological Clinic and Institute for Physical Medicine and Rehabilitation, Stadtspital Triemli, CH-8063 Zürich, Switzerland. Tel: 41-(0)1-466-23-02.; E-mail: robert.theiler@triemli.stzh.ch

Received 23 November 2003; revision accepted 11 August 2004.

a

b

Fig. 1. (a) QUALITOUCH® multimedia 3-D interactive interface—questions answered by touching one of the squares of the Likert scale on the computer screen. (b) Screen display of the electronic touch screen featuring question 6 of the WOMAC 3.1 OA Index and the Likert scale. Translation: Steifigkeit = stiffness. Wie stark ist Ihre Gelenksteifigkeit nach dem ersten Aufwachen am Morgen? = how important is the stiffness of your joints after the first awakening in the morning. Keine = none. Leichte = mild. Mässige = moderate. Starke = severe. Extreme = extreme. Frage wiederholen = repeat question. Befragung abbrechen = stop interview. Hilfe = help. Vorherige Frage = previous question. Nächste Frage = next question.

method). This may be appreciated especially by the older patients. The electronic formats of the WOMAC-Index 3.0 have been previously validated[5–7].

Of special clinical and scientific interest is the ability of a questionnaire to detect and monitor improvement or worsening of the clinical situation based on an intervention. We therefore compared the responsiveness of the original paper format and the computer touch screen format to rofecoxib treatment in patients with symptomatic hip or knee OA over a course of 3 weeks. The aim of this study was to validate the longitudinal responsiveness of the e-WOMAC questionnaire.

Patients and methods

Three sites that participated in the previously published SVIS-Study, with 136 recruited patients and 22 sites in total[5], participated in this ancillary e-WOMAC protocol and 53 eligible consecutive outpatients were recruited and included in the evaluation. The SVIS-study was a prospective open label 3 weeks multicenter study to document the effect of rofecoxib in patients with painful radiographically proven primary OA of the knee or the hip according to ACR criteria who were dissatisfied with their prior NSAID treatment (because of either non-responsiveness to or adverse events from previous NSAID-therapy, including celecoxib). At inclusion the patients stopped their previous NSAID therapy and started therapy with rofecoxib 25 mg once daily on the following day (t_0) for 3 weeks, after which the final visit took place (t_2), with an interim visit on day 7 (t_1)[5].

Because the core study was a clinical trial with drug intervention, conducted according to GCP guidelines and the e-WOMAC is not yet a validated format of the questionnaire acceptable to regulatory authorities, we had to renounce to block randomize the patients for the paper vs the electronic format of the questionnaire. However, all patients gave their separate written informed consent before their participation in this ancillary study. At all three visits, the patients filled in the paper format of the WOMAC first, followed by the electronic format.

The validated German paper format of the WOMAC 3.1 with a Likert scale was used[8,9]. The electronic format of the WOMAC was identical to the German questionnaire with an identical Likert scale in a computerized touch screen format, which has been previously shown to have very good agreement with the original paper format in its numeric rating scale format[6]. The QUALITOUCH® data capture method was developed to facilitate patient assessment. The QUALITOUCH® computer program offers a multimedia 3-D interactive interface: the questions are displayed on a 34.3 cm diameter screen as a text and a situational cartoon and are verbalized over the loudspeaker. The questions are answered by touching one of the squares of the Likert scale on the computer screen [Fig. 1(a)]. By using five buttons on screen the patient could exit the questionnaire, get help, have the question repeated or move only one question forward or backward [Fig. 1(b)]. It is therefore possible to leave out one question and move to the next. Furthermore, the help function self activates after 15 s of inactivity and presents the next possible steps to the patient. Patients are not able to see their prior scores.

STATISTICS

Descriptive statistics included the mean of the aggregated scores, the standard deviation and the mean difference between the scores of the paper and computerized

Table I
Patient demographics

	Male ($n = 32$)	Female ($n = 21$)	Total ($n = 53$)
Age in years (mean ± SD)	63.2 ± 10.3	65.7 ± 8.1	64.2 ± 9.5
Height in cm (mean ± SD)	175.5 ± 5.3	162.0 ± 6.8	170.1 ± 8.9
Weight in kg (mean ± SD)	86 ± 15.3	75.7 ± 17.3	81.9 ± 16.8

Table II
Comparative table of the effect of rofecoxib by format (paper vs electronic) and by WOMAC standardized subscale score over time, mean ± SD

Subscale	Format	t_0	t_1	t_2	Effect size t_2 vs t_0
Pain	Paper	39.7 ± 14.5	33.7 ± 15.7	28.6 ± 14.8	0.76
	e-WOMAC	42.3 ± 15.2	34.6 ± 15.6	29.2 ± 14.6	0.88*
Stiffness	Paper	43.4 ± 18.9	36.5 ± 16.7	31.1 ± 20.0	0.63
	e-WOMAC	46.1 ± 22.2	38.6 ± 18.1	33.4 ± 20.8	0.59*
Function	Paper	44.1 ± 14.0	38.4 ± 15.4	32.8 ± 16.2	0.75
	e-WOMAC	43.8 ± 14.3	38.9 ± 16.2	32.0 ± 16.2	0.77*

*Difference between paper vs electronic format scores statistically not significant.

format. To detect possible format (paper vs computer), time (t_0 = baseline vs t_1 = visit 1 at day 7 vs t_2 = visit 3 at day 21), scale (pain vs stiffness, vs function) or gender (male vs female) related effects, a variance analysis by "repeated measures ANOVA" was performed. In addition, the standardized mean difference was used to measure the effect size by WOMAC subscale between t_0 and t_2 (effect size = [(mean score t_2 − mean score t_0)/(pooled SD)] and tested for significance of the paper vs the electronic format. All statistical tests were performed at a significance level of 0.01 or lower to correct for multiple testing. Normality of the distribution was tested by Kolmogorov–Smirnov. All statistical analyses were performed with SAS® StatView® 5.01.

Results

PATIENTS

All consecutive 53 patients recruited in the three participating study centers were included in the analysis and completed both formats of the questionnaire in all three visits. The detailed patient characteristics are shown in Table I. Age and gender were balanced. Forty-three patients (81%) had primary unilateral knee osteoarthritis, 5 (9%) primary bilateral knee OA, 4 (8%) primary idiopathic hip OA and 1 (2%) had secondary hip OA after congenital hip dysplasia. The time needed to answer all questions in the electronic format of the WOMAC 3.1 was 12.9 ± 2.7 min vs 12.5 ± 3.5 min for the paper format (not significant).

WOMAC-SCORES

WOMAC baseline scores by subscale were not significantly different by format (electronic or paper) and the effect size by subscale between t_0 and t_2 was not significantly different by format (Table II). The overall effect size between t_0 and t_2 was 0.71 for the paper version and 0.74 for the electronic version and was statistically not significant. Therefore, the responsiveness by subscale was not significantly different between the electronic and the paper versions of the WOMAC. At visit 1 (t_1 = day 7) pain and stiffness were significantly reduced with rofecoxib, while function was significantly increased ($P < 0.01$, Table III). The mean magnitude of the effect of rofecoxib was continuously increasing over time in all three subscales and in both formats. Comparing baseline (t_0) to visit 2 (day 21), pain decreased by 30%, stiffness by 28% and function increased by 26% irrespective of whether the paper or the electronic format of the WOMAC 3.1 was used (Fig. 2). While the format (paper vs electronic) might have had some influence on the stiffness scale and on the total WOMAC Index, the format × time interaction of WOMAC 3.1 scores between the paper and the electronic formats at t_0 (baseline), t_1 (day 7) and t_2 (day 21) were not significantly different for pain ($P = 0.22$), stiffness ($P = 0.895$), functional ability ($P = 0.542$) and for the total WOMAC Index ($P = 0.508$), indicating that the pattern of changes in WOMAC scores and subscores over time did not differ by format (Fig. 3). The time × format × scale interaction was not significant, indicating that the pattern of changes of the WOMAC scores was the same for both formats and all scales (Table IV). Gender had no influence on the results.

Discussion

This is the first longitudinal study documenting repeated measures with electronic data capturing through patient self-assessment. Electronic data capturing has become increasingly popular for data acquisition in clinical trials. However, the data collected usually refers to laboratory or diagnostic examination values or to patient history and data entry is usually performed by medical or paramedical personnel. In daily clinical care, there is little experience with patient self-assessment using standardized questionnaires and almost no related validated tools exist. One study documents the initial evaluation of an electronic format of the Short Form 36, concluding that electronic data collection is acceptable to patients and feasible in a clinical setting while providing comparable responses to those of the paper format, improving data capture and being immediately available[10].

Table III
Effect of time (t_0, t_1, t_2), format (paper vs electronic) and time × format interaction on the WOMAC subscales and the WOMAC Index

	t_0, t_1*	t_1, t_2*	t_0, t_2*	Format**	Time × format*
Pain subscale	−6.86 ($P = 0.0002$)	−5.27 ($P = 0.0041$)	−12.14 ($P < 0.0001$)	−1.33 ($P = 0.0119$)	$P = 0.22$
Stiffness subscale	−7.73 ($P = 0.0003$)	−4.77 ($P = 0.0243$)	−12.5 ($P < 0.0001$)	−2.73 ($P = 0.0078$)	$P = 0.895$
Function subscale	5.32 ($P = 0.0018$)	6.25 ($P = 0.0003$)	11.58 ($P < 0.0001$)	0.26 ($P = 0.5770$)	$P = 0.542$
WOMAC 3.1 Index	−6.64 ($P < 0.0001$)	−5.43 ($P = 0.0008$)	−12.07 ($P < 0.0001$)	−1.27 ($P = 0.0049$)	$P = 0.508$

*Significant if $P < 0.0033$. **Significant if $P < 0.01$.

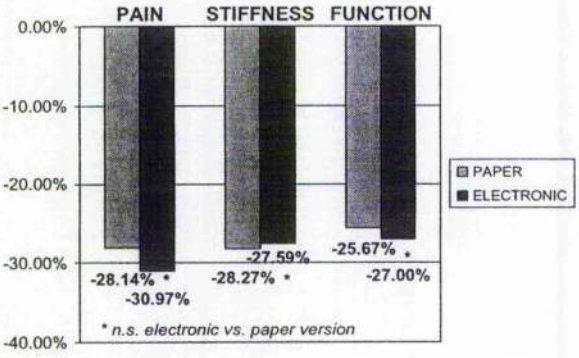


Fig. 2. Evolution of WOMAC 3.1 subscales with rofecoxib one tablet once daily over 3 weeks, paper vs electronic evaluation.

Patient self-assessment by electronic data capturing presents numerous advantages in clinical care: the data collection is standardized and the impact of potential external influences, which may vary in nature from visit to visit, is limited; the data may be collected anonymously across departments, hospitals and medical practices allowing for constant optimization of patient management techniques by detecting outliers regarding treatment success. When performed in the waiting-room, self-assessment makes the best use of the patient's and the physician's time and is a valuable contribution to the patient–physician interaction, especially in the decision-making process of treatment adaptations. In contrast with paper questionnaires which are archived in the patient's file, electronic data allow for easy treatment effect monitoring at a glance. With the e-WOMAC, the patient's progress is documented for the three subscales: pain, stiffness and function. Multiple assessments over time are displayed on one page displaying the change over time in an easy to read graph. The e-WOMAC data collection by patient self-assessment with a QUALITOUCH® touch screen interface has been shown to have very good agreement in all three subscales with the original paper format of the questionnaire[6]. In another study, the patient preference for the electronic vs the paper format of the questionnaire was documented: although 54% of the patients had no experience with computers at all, only 9% preferred the paper format, 91% either preferring the computer format (51%) or being indifferent (38%). Ninety-four percent of the patients declared that the 3-D environment presented (text, sound and cartoon) was helpful[7].

This study shows that e-WOMAC is responsive to treatment over time with regard to pain, stiffness and function. In addition, no significant difference was found while comparing the degree of responsiveness by subscale between the electronic and the paper format of the

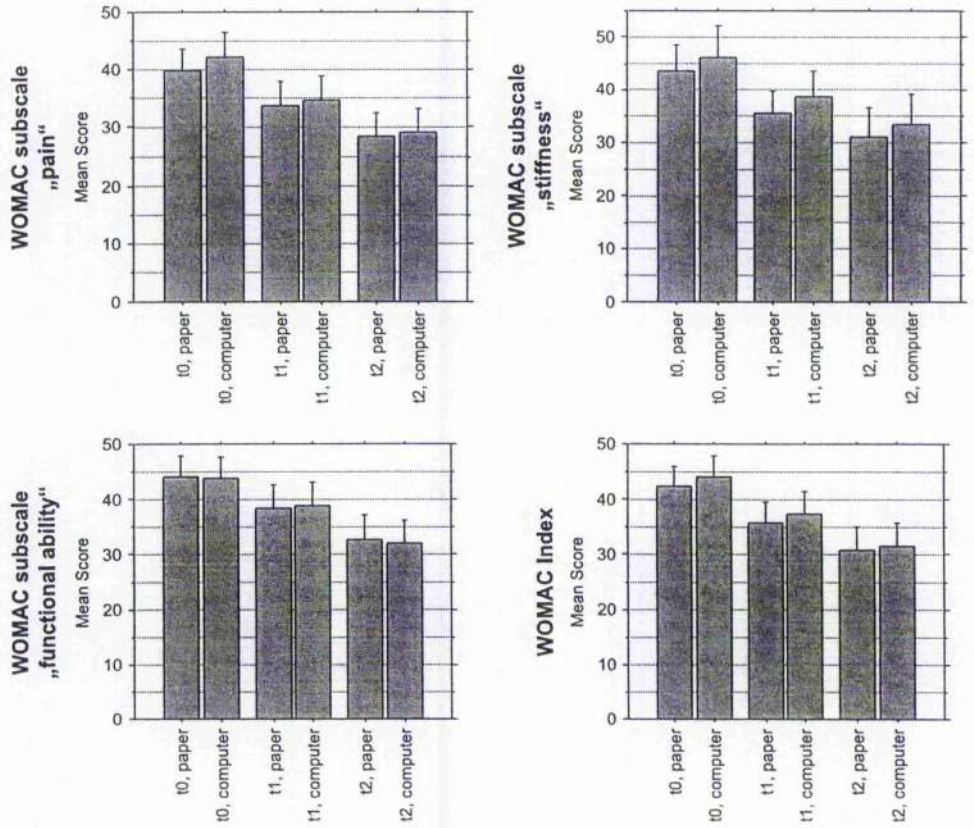


Fig. 3. Changes in mean WOMAC subscales and in WOMAC Index (±95% confidence intervals), by visit at t_0 (baseline), t_1 (day 7) and t_2 (day 21), paper vs computer format, with rofecoxib one tablet once daily.

Table IV
WOMAC Index: levels of significance with repeated measures ANOVA with one, two or three factors

Effect	*P* value	Interpretation
Time* (t_0, t_1, t_2)	<0.0001	Rofecoxib significantly improved WOMAC score over time
Format* (paper vs e-WOMAC)	0.0049	Format had an influence on WOMAC score
Scale* (pain vs stiffness vs function)	0.0183	Subscale had no influence on WOMAC score
Time × format*	0.5077	Changes of WOMAC scores over time were the same for both formats
Time × scale*	0.7229	Changes of WOMAC scores over time were the same for all WOMAC subscales
Time × format × scale*	0.7703	Patterns of change in WOMAC scores over time did not differ by version and scale

*Significant if $P < 0.01$.

questionnaire. This suggests that the e-WOMAC is as responsive as the original paper format. As the main endpoints of the SVIS study were to document the effects of rofecoxib on Quality of Life (measured by the SF-12) and disease specific symptoms (measured by the WOMAC paper questionnaire) and because at the time of the initiation of the SVIS study the e-WOMAC was not completely validated, we renounced to randomize for the two formats (paper and electronic) and asked all patients to fill in the paper format of the questionnaire consistently before they filled in the electronic format, accepting thereby a systematic error in the validation procedure. As a single parameter, the format seemed to have an influence on the WOMAC score, the significance being driven by the two questions related to pain (Table V). In this study, the overall effect size between t_0 and t_2 reached 0.71 when measured with the paper version of the WOMAC and 0.74 with the electronic version. An effect size between 0.2 and 0.5 is considered as small but clinically meaningful, while a large effect size is estimated at being 1.0 or more[11]. Therefore the observed effect size of rofecoxib between t_0 and t_2 should be considered not only statistically significant but also clinically relevant. In contrast the difference in effect size of 0.03 observed between the paper and the electronic version of the WOMAC is statistically non-significant and should be considered as clinically irrelevant. This holds true for all three subscales of the WOMAC, the largest observed difference in effect size between formats being 0.12. Furthermore, the paper and the electronic formats of the WOMAC have proven to be very similar for the monitoring of treatment effects and under the premises that the choice for the paper or the electronic format is made upfront and carried out throughout the timecourse of the observation, both formats can be considered as equivalent. In the meantime, another study with correct block randomization has demonstrated the patient's preference for the electronic format and the perfect interchangeability of the paper and the electronic formats[7].

Table V
Influence of the version in relation with the section of the questionnaire

Questions addressing	Number of questions	Significance level
Pain	5	0.012
Stiffness	2	0.0078
Function	17	0.577

*Significant if $P < 0.01$.

In patients with symptomatic OA at the hip or knee treated with rofecoxib, the paper and the electronic format of the WOMAC 3.1 showed similar effect sizes and were equally suitable for the longitudinal monitoring of the effects of drug treatment and the detection of clinically meaningful changes. The future successful use of the e-WOMAC by the QUALITOUCH® method in medical care will depend on its integration in the daily processes of patient management at the primary care physician level.

References

1. Bellamy N, Buchanan WW. A preliminary evaluation of the dimensionality and clinical importance of pain and disability in osteoarthritis of the hip and knee. Clin Rheumatol 1986;5(2):231–41.
2. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988;15(12):1833–40.
3. Bellamy N, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P, *et al.* Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. J Rheumatol 1997;24(4):799–802.
4. Pincus T, Sokka T. Quantitative measures for assessing rheumatoid arthritis in clinical trials and clinical care. Best Pract Res Clin Rheumatol 2003;17(5):753–81.
5. Theiler R, Bischoff HA, Good M, Uebelhart D. Rofecoxib improves quality of life in patients with hip or knee osteoarthritis. Swiss Med Wkly 2002;132(39–40):566–73.
6. Theiler R, Spielberger J, Bischoff HA, Bellamy N, Huber J, Kroesen S. Clinical evaluation of the WOMAC 3.0 OA Index in numeric rating scale format using a computerized touch screen version. Osteoarthritis Cartilage 2002;10(6):479–81.
7. Bischoff-Ferrari HA, VonDechend M, Bellamy N, Theiler R. Validation and patient acceptance of a computer touch screen version of the WOMAC 3.1 Osteoarthritis Index. Ann Rheum Dis July 1 2004; PMID:15231508.
8. Stucki G, Meier D, Stucki S, Michel BA, Tyndall AG, Dick W, *et al.* [Evaluation of a German version of WOMAC (Western Ontario and McMaster Universities) Arthrosis Index]. Z Rheumatol 1996;55(1):40–9.
9. Bellamy N. Personal communication of the German version of the WOMAC 3.1 with Likert scale.
10. Wilson AS, Kitas GD, Carruthers DM, Reay C, Skan J, Harris S, *et al.* Computerized information-gathering in specialist rheumatology clinics: an initial evaluation of an electronic version of the Short Form 36. Rheumatology (Oxford) 2002;41(3):268–73.
11. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd edn. Hillsdale, NJ: L. Erlbaum Associates; 1988: p. xxi, 567.

Current Medical Research and Opinion, Vol. 15, No. 2, 1999, 113–119

Comparative Study of Self-rating Pain Scales in Osteoarthritis Patients

Nicholas Bellamy[1], Jane Campbell[2] and Jerry Syrotuik[3]

[1]*University of Queensland, Brisbane, Australia*
[2]*University of Western Ontario, London, Canada*
[3]*Searle Canada, Mississauga, Canada*

Key words: *Osteoarthritis – Pain – Outcome measurement*

Summary

Although progress has been made in the clinical metrology of pain in osteoarthritis, much further work remains. The preferred methods of measurement remain debatable. In this longitudinal, open study, a comparison of eight self-rating pain scales has been conducted. A total of 333 patients entered the four-week study after completing a 3–7 day NSAID-free washout period. Patients were assigned to treatment with oxaprozin 1200 mg p.o. once daily with titration permitted between 600 mg and 1800 mg. Rescue analgesia with acetaminophen (paracetamol) 325 mg (maximum 2600 mg) was allowed. At the end of the washout and the treatment period, patients completed eight self-administered pain scales.

All pain measures detected clinically important and statistically significant improvements in pain. The pain scales differed in their degree of responsiveness. The Likert and visual analogue scales and their primary variations (continuous chromatic analogue and numerical scales) were more responsive than more complex measures. A positive correlation between initial pain rating and subsequent pain relief was confirmed in this study.

We conclude that, while pain is a subjective sensory phenomenon, its perceived severity can be evaluated using a variety of self-administered pain scales, all of which are capable of detecting improvements in health status following effective pharmacological intervention.

Introduction

Pain is an entirely subjective phenomenon and is the quintessential symptom of most musculoskeletal conditions. The methods used to assess pain are many and varied. With the exception of the behavioural observation techniques, most methods are based on patient self-report questionnaires[1]. Many of the most sophisticated health status questionnaires currently available contain pain questions or even distinct pain subscales. It is debatable whether it is preferable to measure pain as a global entity or to measure pain in each of several distinct situations (e.g. night pain, pain with activity, pain at rest), and then create a pain subscale which respects the totality of pain. The clinimetric problem is further complicated by the fact that pain may be different in different joints at the same point in time, and, in a single joint,

Address for correspondence: Prof. Nicholas Bellamy, Director – Centre of National Research and Rehabilitation Medicine (CONROD), C Floor, Clinical Sciences Building, University of Queensland Department of Medicine, Royal Brisbane Hospital, Herston, Brisbane, Queensland 4029, Australia
Accepted: 22nd February 1999

and may fluctuate with time, sometimes, but not always, in a rhythmic circadian or circaseptan pattern[2]. In selecting a pain scale for longitudinal monitoring of patients with osteoarthritis (OA), there are at least two important issues:

1. For clinical practice and clinical trials purposes, there are differences in the responsiveness (syn: sensitivity to change) of different pain scales, and
2. In clinical trials, should pain be used as a stratification variable because of a purported relationship between current pain severity and the magnitude of the subsequent pain relief achieved with treatment?

Although a few comparative studies of pain rating scales in musculoskeletal diseases have been conducted, the diversity of scales is now greater, and, therefore, it is timely to reappraise this important issue, based on the principal types of scales employed currently in the evaluation of OA patients.

Patients and Methods

The study was conducted as a longitudinal open study in clinical practice settings including general practitioners, rheumatologists and internists. Patients were required to fulfil American College of Rheumatology criteria for the diagnosis of OA[3], be aged 18–75 years, ambulatory, able and willing to complete the study questionnaires, able and willing to give informed consent, and be eligible to be treated with a non-steroidal anti-inflammatory class agent for at least four weeks. Patients were excluded for the following reasons:

1. Pregnant or nursing mothers
2. Applying for, or receiving, disability or workmen's compensation benefits
3. Currently engaged in disability-related litigation
4. Allergy to aspirin or other NSAIDs
5. Allergy to, or intolerance of, acetaminophen
6. Active gastritis currently, or documented peptic ulcer disease in the last six months
7. Upper gastrointestinal bleeding in the last two years
8. Uncontrolled hypertension, respiratory, renal, hepatic, gastrointestinal, haematological, endocrine or any other disease which, in the opinion of the investigator, might affect the evaluation of the study medication
9. Uncontrolled cardiac disease
10. Previously entered in the protocol
11. Recipient of any investigational drug 30 days before study entry
12. Requiring concomitant treatment with narcotic analgesics
13. Systemic or intra-articular steroid injections or soft tissue injection of a steroid in the last month
14. Not practising a reliable method of contraception
15. Concomitant treatment with anticoagulant or lithium
16. Any inflammatory disease other than rheumatoid or osteoarthritis
17. Evidence of chondrocalcinosis on available radiograph

Consenting patients completed a 3–7-day NSAID-free washout period and were thereafter placed on oxaprozin (Daypro™) 1200 mg p.o. once daily. Bidirectional dose titration of between 600 mg and 1800 mg per day was permitted during the active treatment phase. The medication was taken as a single dose except at the 1800 mg dose, where the medication was split between 1200 mg in the morning and 600 mg in the evening. All medications were taken with food. The planned active treatment phase

was four weeks. During the washout phase and active treatment period, rescue analgesia with acetaminophen (paracetamol) 325 mg (maximum eight tablets a day in divided dosage) was permitted. Compliance to study medication was assessed by pill counting. Patients were asked to avoid analgesia for eight hours prior to the baseline assessment.

In addition to collecting demographic and disease data, patients completed the following self-administered pain rating scales at the end of the washout period and again at the end of the study: McGill pain questionnaire (MPQ)[4], five-point Likert scale[1], 10 cm horizontal visual analogue scale[1], ladder scale (reversed)[5], numerical rating scale[1], continuous chromatic analogue scale[6], Moll pain faces scale[7] and Champion pain faces scale[8]. Both English and French-Canadian versions of the pain scales were available since the study was run in various centres across Canada. The sample size was considered sufficient to provide a broad experience with the various pain scales. Study subjects were characterised using descriptive statistics. The mean and standard deviation (SD) of the baseline scores, termination scores and change scores were calculated. The static and change scores have been documented to facilitate the calculation of sample size for future clinical studies based on the parameters observed in this study. The relative responsiveness of the instruments was compared using *t*-values[9] and standardised response means (SRMs)[10]. The five-point Likert scale was selected arbitrarily as the anchor for all RE comparisons. Within-group changes were assessed using the Student's *t*-test based on a per protocol analysis. Finally, the relationship between the initial pain rating and the subsequent pain response (pain relief) was determined using Pearson correlation coefficients. Since the efficacy of oxaprozin in OA had been previously established[11–13], we were interested in comparing the responsiveness of different pain scales, and, therefore, performed a per protocol (efficacy) analysis rather than an intention to treat (effectiveness) analysis.

Results

Of the 333 OA patients who entered the study, 119 were excluded from analysis for the following reasons:

1. Age > 75 years (14)
2. Compliance < 80% (28)
3. Use of other analgesics (5)
4. Assessments completed outside study duration window of ± 5 half-lives (i.e. 10 days) (72)

Response data from 214 patients were analysed, of whom 61% were females and 92% were Caucasian. The mean age of participants was 60 years (range 35–83 years, SD = 11) and mean disease duration was seven years (range 0.1–40 years, SD = 7). The mean joint count was 5 (range 1–41, SD = 7). Mean pain scores and SD at baseline and termination, change (relief) scores, SD, *p*-values and effect sizes are shown in Table 1. Clinically important and statistically significant treatment effects were recorded by all eight pain scales, including the different components of the MPQ. Pain measures have been ranked according to the hierarchy of effect sizes. Statistical *p*-values were similar for most measures. In contrast, *t*-values and SRMs were different for different measures (Table 1). The most commonly chosen words at baseline and termination in the English and French-Canadian version of the MPQ are illustrated in Table 2. The hierarchy of words chosen was slightly different in English versus French-Canadian centres and differed at termination versus baseline.

Table 1. Mean pain scores at baseline, termination and relief score ordered by effect size

Measure	Baseline		Termination		Change (relief) score			SRM
	Mean	SD	Mean	SD	Mean	SD	*p*-value	
VA	59.71	21.78	35.45	24.17	24.73	25.59	0.0001	0.97
CCAS	5.77	1.93	3.66	2.27	2.16	2.31	0.0001	0.94
Numerical	5.86	2.00	3.70	2.25	2.17	2.38	0.0001	0.91
Likert	3.38	0.72	2.60	0.85	0.79	0.94	0.0001	0.84
Pain faces 1[8]	3.39	1.44	1.99	1.60	1.43	1.84	0.0001	0.78
MPQ (total)	24.50	12.78	16.14	12.44	8.79	11.87	0.0001	0.74
MPQ (PPI)	2.70	0.90	1.99	0.92	0.71	1.02	0.0001	0.70
MPQ (sensory)	15.49	7.52	10.78	7.30	5.17	7.40	0.0001	0.70
Pain faces 2[7]	4.47	1.19	3.37	1.64	1.11	1.59	0.0001	0.70
MPQ (NWC)	10.39	4.83	7.63	5.38	2.87	4.40	0.0001	0.65
Reversed ladder	5.40	2.04	4.03	2.55	1.38	2.79	0.0001	0.50
MPQ (misc.)	5.03	3.20	4.28	2.99	1.45	3.28	0.0001	0.44
MPQ (evaluative)	2.49	1.21	2.02	1.13	0.55	1.42	0.0001	0.39
MPQ (affective)	3.86	2.79	3.29	2.67	0.91	2.80	0.0020	0.33

Key: SRM, standardised response mean; VA, visual analogue; CCAS, continuous chromatic analogue scale, MPQ, McGill pain questionnaire; PPI, present pain intensity; NWC, number of words chosen; Misc, miscellaneous.

Table 2. Words chosen by patients from the McGill pain questionnaire to describe their pain by language of the questionnaire

English				French			
Baseline $n = 183$		Termination $n = 183$		Baseline $n = 31$		Termination $n = 31$	
Word	%	Word	%	Word	%	Word	%
Nagging	52	Aching	35	Elancement	52	Fatigante	39
Aching	51	Tender	35	Fatigante	42	Elancement	26
Throbbing	49	Tiring	34	Desagreable	35	Engourdissement	26
Shooting	40	Annoying	31	Enervante	29	Enervante	23
Sharp	38	Nagging	30	Engourdissement	29	Desagreable	19
Tiring	37	Burning	20	Serrement	29	Harcelante	19
Tender	36	Shooting	19	Chaleur	26	Chaleur	16
Burning	30	Throbbing	18	Deprimante	26	Deprimante	10
Exhausting	30	Sharp	18	Harcelante	26	Serrement	10
Miserable	28	Miserable	18	–	–	–	–
Stabbing	25	Penetrating	16	–	–	–	–
Gnawing	25	Gnawing	15	–	–	–	–
Penetrating	25	Stabbing	12	–	–	–	–
Annoying	25	Exhausting	10	–	–	–	–

Statistically significant correlations were observed between baseline scores and subsequent pain relief scores (Table 3).

Discussion

The measurement of pain in clinical trials requires the use of valid, reliable and responsive assessment techniques. The clinimetric properties of scales used in the study have been previously evaluated[1]. Previous comparative studies of pain scales have suggested high levels of correlation between scores[7,14]. However, correlation coefficients are an expression of the degree of association, not measures of agreement, nor do they address issues of the relative responsiveness of different scales. Data derived from a study of auranofin in RA suggested that, while scale sensitivity was comparable for different measures, the *p*-value for the 10 cm VA scale ($p = 0.01$) was

Table 3. Correlations of initial pain rating and pain relief

Pain scale	*r*	*p*-value
MPQ (evaluative)	0.66	0.0001
MPQ (miscellaneous)	0.58	0.0001
MPQ (affective)	0.56	0.0001
Pain faces 1[6]	0.55	0.0001
MPQ (present pain intensity)	0.54	0.0001
MPQ (sensory)	0.52	0.0001
Likert	0.50	0.0001
MPQ (total score)	0.50	0.0001
Visual analogue scale	0.49	0.0001
Numerical point	0.48	0.0001
Ladder	0.48	0.0001
Continuous chromatic analogue scale	0.43	0.0001
Pain faces 2[7]	0.34	0.0297
MPQ (number of words chosen)	0.32	0.0001

MPQ, McGill pain questionnaire

less than that for the MPQ ($p = 0.02$), and both were smaller than for the pain ladder scale ($p = 0.09$)[5]. Previous studies have either focused on patients with osteoarthritis[7], or mixed musculoskeletal disorders[14], or have compared relatively few pain measures in RA[5]. By comparison, we have assessed a large number of different scales in a single well-defined disorder. We have observed that all measures detected clinically important and statistically significant improvements in pain during treatment with oxaprozin. The clinical efficacy of Daypro™ in OA has been previously established in randomised clinical trials[11–13] using a variety of pain scales. We can now confirm that efficacy, regardless of the type of pain scale employed. Pain scales, however, do differ in their degree of responsiveness. It is of note that the two basic approaches to pain measurement (VA and Likert scales) and their primary variation (CCAS and numerical scales) were more responsive than more complex measures. In particular, we were not able to demonstrate any statistical superiority of the MPQ over other indices. The MPQ is a more complex index which does indeed provide qualitative, as well as quantitative information, regarding the patient's pain. It is of note that French- and English-speaking Canadians used different words from one another, and this is probably culturally based, since there is no evidence that the disease is differently expressed in different parts of Canada. For clinical trials purposes, the most responsive pain measure would afford a reduction in sample size. It appears, therefore, that the VA and Likert scales frequently used are indeed appropriate for this purpose. However, the CCAS is difficult to produce in comparison with other measures and may not be quite as practical, and the complexities of pain faces scales, while offering advantages in special subgroups (e.g. children), may not be accompanied by superior responsiveness. In contrast, numerical rating scales may be useful, particularly in transcultural adaptations of pain questionnaires given the commonality of numerical (cf. linguistic) expression in different countries.

By comparison, in clinical practice, the key requirements of a measurement are simplicity, brevity, rapid completion and ease of scoring[15,16]. The LK and numerical scales are likely to be the most useful in this setting. Quantitative measurement, using standardised self-reported health-status measures, is performed relatively infrequently in rheumatology outpatient practice, according to surveys conducted recently in Canada[15] and Australia[16].

Studies, examining the contribution of serial quantitative clinical measurement to clinical decision making and patient outcome, are urgently required to define the role of quantitative measurement in routine practice.

The positive correlation between initial pain rating and subsequent pain response, as measured using a subtraction technique ($r = 0.62$) originally noted by Huskisson[17], and confirmed in this study, suggests that patients with more severe pain may achieve greater overall reductions in pain than those with mild pain. It should, however, be noted that, despite being statistically significant, the strength of association does vary quite markedly for different pain scales (0.32–0.66). These observations have important implications for clinical trials where initial pain rating might be used as a stratification variable, and for some pain scales may have more relevance than demographic-based stratification variables such as age and gender or disease-based variables such as disease duration. Although the occurrence of ceiling effects cannot be excluded by this analysis, these observations provide a basis for optimism when treating OA patients with efficacious non-steroidal anti-inflammatory drugs such as oxaprozin, since there appears to be a positive relationship between current pain severity and the degree of pain release subsequently achieved.

We conclude that, while pain is a subjective sensory phenomenon, its perceived severity can be evaluated using a variety of self-administered pain scales, all of which are capable of detecting clinically important, statistically significant improvements in health status following effective pharmacological intervention. Simple scaling methods seem to be as, or more, responsive than more complex methods. Such scales can be incorporated into single-item or multi-item pain questionnaires depending on whether a single global estimate is required or a multi-faceted situationally based estimate is needed, and are equally applicable in the clinical research or clinical practice setting.

Acknowledgements

We wish to thank those who provided instruments: C. Bombardier (ladder scale), G. D. Champion (faces scale), R. Melzack (McGill pain questionnaire), J. M. H. Moll (faces scale) and McNeil Consumer Products (continuous chromatic analogue scale). Data entry and statistical analysis was completed by Innovus Research Inc. This work was supported by Searle Canada.

References

1. Bellamy, N. (1993). Pain measurement. *Musculoskeletal Clinical Metrology*. Kluwer Academic Publishers, Dordrecht, pp. 65–76.
2. Bellamy, N., Sothern, R. B., Campbell, J. and Buchanan, W. W. (1991). Circadian rhythm in pain, stiffness, and manual dexterity in rheumatoid arthritis: Relation between discomfort and disability. *Ann. Rheum. Dis.*, **50**, 243–248.
3. Altman, R. D. (1991). Criteria for classification of clinical osteoarthritis. *J. Rheumatol.*, **18** (Suppl 27), 10–12.
4. Melzack, R. and Katz, J. (1992). The McGill pain questionnaire: Appraisal and current status. In: *Handbook of Pain Assessment* (Eds Turk, D. C. and Melzack, R.). Guilford Press, New York, pp. 152–168.
5. Bombardier, C., Ware, J., Russell, I. J., *et al.* (1986). Auranofin therapy and quality of life in patients with rheumatoid arthritis. Results of a multicenter trial. *Amer. J. Med.*, **81**, 565–578.
6. Grossi, E., Borghi, C., Cerchiari, B. L., Della Puppa, T. and Francucci, B. (1983). Analogue chromatic continuous scale (ACCS): a new method for pain assessment. *Clin. Exp. Rheumatol.*, **1**, 337–340.
7. Frank, A. J. M., Moll, J. M. H. and Hort, J. F. (1982). A comparison of three ways of

measuring pain. *Rheumatol. Rehabil.*, **21**, 211–217.
8. Bieri, D., Reeve, R., Champion, G. D., Addicoat, L. and Ziegler, J. B. (1990). The Faces Pain Scale for the self-assessment of the severity of pain experienced by children: development, initial validation, preliminary investigation for ratio scale properties. *Pain*, **41**, 139–150.
9. Armitage, P. and Berry, G. (1994). Statistical inference. *Statistical Methods in Medical Research*. Blackwell Science, Oxford, 3rd edn, pp. 93–153.
10. Norman, G. R. and Streiner, D. L. (1994). Comparing two groups – the *t*-test. *Biostatistics – The Bare Essentials*. Mosby, St. Louis, pp. 58–62.
11. Waxman, J. (1993). The efficacy and safety of oxaprozin versus piroxicam: Pooled results of double-blind trials in osteoarthritis. *Drug Therapy*, **23** (March Suppl), 13–16.
12. Weaver, A., Rubin, B., Caldwell, J., *et al.* (1995). Comparison of the efficacy and safety of oxaprozin and nabumetone in the treatment of patients with osteoarthritis of the knee. *Clin. Ther.*, **17**(4), 735–745.
13. Makarowski, W., Weaver, A., Rubin, B. *et al.* (1996). The efficacy, tolerability, and safety of 1200 mg/d of oxaprozin and 1500 mg/d of nabumetone in the treatment of patients with osteoarthritis of the knee. *Clin. Ther.*, **18**(1), 114–124.
14. Downie, W. W., Leatham, P. A., Rhind, V. M., Wright, V., Branco, J. A. and Anderson, J. A. (1978). Studies with pain rating scales. *Ann. Rheum. Dis.*, **37**, 378–381.
15. Bellamy, N., Kaloni, S., Pope, J., Coulter, K. and Campbell, J. (1998). Quantitative rheumatology: A survey of outcome measurement procedures in routine rheumatology outpatient practice in Canada. *J. Rheumatol.*, **25**, 852–858.
16. Bellamy, N., Muirden, K., Brooks, P. M., Barraclough, D., Tellus, M. M. and Campbell, J. (1998). A survey of outcome measurement procedures in routine rheumatology outpatient practice in Australia. *J. Rheumatol.* Accepted for publication.
17. Huskisson, E. C. (1974). Measurement of pain. *The Lancet*, **2**, Nov. 9, 1127–1131.

Quantitative Rheumatology: A Survey of Outcome Measurement Procedures in Routine Rheumatology Outpatient Practice in Canada

NICHOLAS BELLAMY, SHAILA KALONI, JANET POPE, KEITH COULTER, and JANE CAMPBELL

ABSTRACT. ***Objective.*** To assess the extent to which quantitative clinical measurement is performed by rheumatologists in the longitudinal followup of patients with rheumatoid arthritis (RA), osteoarthritis (OA), ankylosing spondylitis (AS), and fibromyalgia (FM) in routine outpatient practice in Canada.

Methods. A cross sectional postal survey was conducted using an 18 item self-administered questionnaire sent to Canadian Rheumatology Association members.

Results. Rheumatologists (response rate 85%) were more likely to longitudinally follow patients with RA and AS than those with OA or FM. There was a high degree of variability in the methods used to monitor patients longitudinally. Many measures used in clinical research were used infrequently in routine clinical practice. In general, the major health status measures surveyed were not used in clinical monitoring. There was a high level of agreement (>80%) that the characteristics required of an outcome measure for use in clinical practice should include simplicity, brevity, ease of scoring, reliability, validity, and sensitivity to change.

Conclusion. The majority of Canadian rheumatologists perform outcome measurement during the longitudinal followup of their outpatients with RA, AS, OA, and FM. However, the process lacks standardization. High performance health status measures, developed for clinical research, have not been widely adopted in rheumatology practices. There is agreement on the characteristics required by Canadian rheumatologists for measurement procedures used in routine clinical care. Quantitative measurement in clinical practice using standardized procedures is an attainable, but as yet, unrealized opportunity. (J Rheumatol 1998;25:852–8)

Key Indexing Terms:
OUTCOME MEASUREMENT CLINICAL PRACTICE RHEUMATIC DISEASES

There has been steady progress in the development of measurement techniques for clinical research purposes[1]. Accepted methodologies have been established for the development and validation of new measurement procedures. This evolution has resulted in the availability of a wide variety of outcome measurement alternatives for musculoskeletal clinical trials. One consequence of this development has been a lack of adequate standardization, different measures often being used in different studies[1]. Even regarding primary outcomes, international agreement has been reached only recently on core sets of measures for future Phase III clinical trials in rheumatoid arthritis (RA)[2], and hip, knee, and hand osteoarthritis (OA)[3]. Various groups are working toward developing consensus on core sets for ankylosing spondylitis (AS) and fibromyalgia (FM) studies.

In contrast, there have been no published studies and few recommendations regarding outcome measures for routine clinical care in rheumatology. Furthermore, few techniques have been developed for specific application in the clinical practice setting. In view of these developments, we surveyed monitoring practices used by Canadian rheumatologists in the longitudinal followup of patients with RA, OA (hip, hand, knee, generalized), AS, and FM in routine clinical care. Our purpose was to describe current monitoring practices, to determine the required characteristics of instruments suitable for use in clinical practice, and to gauge the extent to which several major health status instruments are currently being used in the clinical care setting.

MATERIALS AND METHODS

An 18 item (178 subcomponent) questionnaire was developed, pre-tested, revised, formatted, and distributed by post to Canadian rheumatologists. The survey was termed the Outcome Measurement In Rheumatology Routine Outpatient Practice (OMIRROP) Survey. The sample was ascertained from the 1995 Canadian Rheumatology Association (CRA) Directory. We excluded any identified CRA registrant who was not a clinical rheumatologist (e.g., immunologist), or who was a pediatric rheuma-

From the Department of Medicine, University of Western Ontario, London, Ontario, Canada.

N. Bellamy, MD, Msc, FRCP(Glas), FRCP(Edin), FRCPC, FACP, Professor of Medicine; S. Kaloni, MD, Rheumatology Fellow; J. Pope, MD, MPH, FRCPC, Assistant Professor of Medicine; K. Coulter, MD, FRCPC, Associate Professor of Medicine; J. Campbell, BA, Research Assistant.

Address reprint requests to Dr. N. Bellamy, Department of Rheumatology, London Health Sciences Centre-Victoria Campus, Suite 303, Colbourne 3, 375 South Street, London, Ontario N6A 4G5, Canada.

Submitted July 15, 1997 revision accepted November 19, 1997.

tologist, or who was not currently residing in Canada. Two hundred fifty eligible practising rheumatologists were surveyed. Participants were questioned specifically regarding their measurement practices in the longitudinal followup (serial assessments over time) of their adult outpatients with RA, OA, AS, and FM. Because of the large number of outcome measures currently available, it was not possible to include all in the questionnaire. However, the majority were incorporated. For the purpose of the survey, an outpatient was defined as a non-hospitalized (i.e., ambulatory) patient seen either in private clinical practice or in the outpatient clinic of a health care facility. Second and third mailings of the OMERROP questionnaire were made to non-respondents at intervals of about one month, with a personal letter accompanying the third mailing to maximize the response rate.

The analysis was based mainly on descriptive statistics. To identify measurement procedures routinely used by a high percentage of respondents, we separately identified those outcome measures that were used "Always" or "Usually" by ≥ 70% of respondents. While this was an arbitrary division, it nevertheless defines those measures that might be considered part of usual care.

RESULTS

Response data. Responses were obtained from 213 rheumatologists (response rate 85%). The mean year of graduation from medical school of respondents was 1972 (range 1942–1990) (non-respondents: mean 1975, range 1947–1990; p = NS), and the mean year of starting practice in rheumatology of respondents was 1980 (range 1952–1995). The type of practice of respondents was as follows: full time private practice 41%, full time university 34%, part time university 24%. The majority of respondents (84%) had experience participating in at least one prior clinical research project, in which they had been required to make or supervise clinical measurements on study subjects. Respondents were more likely to longitudinally follow patients with RA (100%) and AS (96%) than patients with knee OA (74%), generalized OA (72%), hip OA (71%), hand OA (58%), or FM (51%).

Basic measurement procedures. To assess the use of basic measurement procedures, participants were asked to respond to a number of questions in the following format. "How often do you *serially* use the following assessment techniques for *longitudinally* monitoring the efficacy of antirheumatic drug therapy in your adult (*specify disease*) outpatient practice?" Responses to this question were separately obtained with respect to RA (Table 1), generalized OA (Table 2), hip OA, knee OA, and hand OA (Table 3), AS (Table 4), and FM (Table 5). For RA, participants also were questioned regarding monitoring practices in the separate situations of nonsteroidal antiinflammatory drug (NSAID), disease modifying antirheumatic drug (DMARD), and steroid therapy (Table 1).

Usage patterns for RA, OA, AS, and FM varied for different outcome measures (Tables 1–5). Those used Usually or Always by ≥70% are identified by an asterisk and those used by ≤20% have been relegated to a footnote (Tables 1–5). Measures have been ranked according to the prevalence of Always/Usually usage (Tables 1,2,4,5).

Health status instruments. No major health status instrument evaluated was used frequently in routine clinical practice. The American College of Rheumatology (ACR) Functional Classification[4] was used most frequently (RA 49%, AS 30%, OA 24%, FM 15%), followed by the Health Assessment Questionnaire (HAQ)[5] (RA 16%, AS 11%, OA 9%, FM 9%), the Functional Status Index[6] (RA 12%, AS 9%, OA 7%, FM 7%), and the Fibromyalgia Impact Questionnaire[7] (FM 8%). For all other instruments, i.e.,

Table 1. Responses to the questions: (A) How often do you serially use the following assessment techniques for longitudinally monitoring the efficacy of antirheumatic drug therapy in your adult RA outpatient practice? (B) In what situation do you use the aforementioned assessment techniques in RA? (please only answer for those techniques that you do use).

Assessment Technique	A				B		
	Nev	Occ	Usl	Alw	Monitoring NSAID Therapy	Monitoring DMARD Therapy	Monitoring Steroid Therapy
Duration of morning stiffness*	2	5	31	62	82	94	80
Physician global assessment (same/better/worse)*	7	6	31	56	81	91	81
Patient global assessment (same/better/worse)*	6	8	37	49	80	88	76
Number of swollen joints*	3	13	38	46	71	89	72
Number of tender joints*	5	13	40	42	73	87	73
Number of involved joints*	10	15	37	38	59	77	60
Physician global assessment (none/mild/moderate/severe)	21	13	34	32	68	76	66
Number of damaged joints	12	29	34	25	40	68	48
Severity of morning stiffness	23	20	27	30	60	69	60
Patient global assessment (none/mild/moderate/severe)	29	20	26	25	44	71	61
Grip strength	19	34	26	21	53	64	53
Pain scale (adjectival)	46	19	25	10	64	67	59
ARA joint count	53	20	13	14	41	54	39

Nev: Never, Occ: Occasionally, Usl: Usually, Alw: Always.
*Used by ≥ 70% of respondents Usually or Always.
NB: Used by ≤ 20% of respondents Usually or Always: Pain scale (VAS), Physician global assessment (VAS), Patient global assessment (VAS), 28 joint count, Some other form of joint count (specify), Walk time, Ritchie articular index, Some other form of pain scale (specify).

Table 2. Responses to the question: How often do you serially use the following assessment techniques for longitudinally monitoring the efficacy of antirheumatic drug therapy in adult outpatients with generalized OA?

Assessment Technique	Never	Occasionally	Usually	Always
Patient global assessment (same/better/worse)*	14	6	40	40
Physician global assessment (same/better/worse)*	15	5	40	40
Number of involved joints	24	15	32	29
Physician global assessment (none/mild/moderate/severe)	30	9	31	30
Number of swollen joints	25	16	31	28
Patient global assessment (none/mild/moderate/severe)	31	10	29	30
Number of damaged joints	24	20	30	26
Number of tender joints	26	21	27	26
Pain scale (adjectival)	39	11	30	20
Duration of morning stiffness	28	24	28	20
Severity of morning stiffness	45	22	19	14
Grip strength	44	31	13	12

*Used by ≥ 70% of respondents Usually or Always.
NB: Used by ≤ 20% of respondents Usually or Always: Pain scale (VAS), Patient global assessment (VAS), Walk time, Physician global assessment (VAS), ARA joint count, Some other form of pain scale (specify), Some other form of joint count (specify), 28 joint count, Doyle articular index.

Table 3. Responses to the question: How often do you serially use the following assessment techniques for longitudinally monitoring the efficacy of antirheumatic drug therapy in your adult OA outpatient practice in patients with (A) Hand OA, (B) Knee OA, and (C) Hip OA?

Assessment Technique	(A) Hand OA				(B) Knee OA				(C) Hip OA			
	Nev	Occ	Usl	Alw	Nev	Occ	Usl	Alw	Nev	Occ	Usl	Alw
Functional capacity*	13	5	41	41	10	2	37	51	9	4	37	50
Pain scale (VAS or adjectival)	37	15	28	20	27	14	33	26	27	11	31	31
Duration of morning stiffness	27	24	30	19	27	27	28	18	29	24	28	19
Severity of morning stiffness	42	26	20	12	42	23	23	12	39	25	23	13
Patient global assessment*	18	7	34	41	—	—	—	—	—	—	—	—
Physician global assessment*	19	8	34	39	—	—	—	—	—	—	—	—
Number of involved joints	23	11	34	32	—	—	—	—	—	—	—	—
Number of tender joints	24	15	33	28	—	—	—	—	—	—	—	—
Number of swollen joints	27	14	32	27	—	—	—	—	—	—	—	—
Grip strength	34	27	22	17	—	—	—	—	—	—	—	—
Swelling (present/absent)*	—	—	—	—	8	5	35	52	—	—	—	—
Crepitus (present/absent)*	—	—	—	—	10	9	31	50	—	—	—	—
Tenderness (present/absent)*	—	—	—	—	11	10	37	42	—	—	—	—
Knee ROM† (by goniometry)	—	—	—	—	36	21	22	21	—	—	—	—
Tenderness (graded)	—	—	—	—	41	23	21	15	—	—	—	—
Hip ROM (by goniometry)	—	—	—	—	—	—	—	—	43	15	23	19

Nev: Never, Occ: Occasionally, Usl: Usually, Alw: Always, ROM: range of motion.
*Used by ≥ 70% of respondents Usually or Always.
NB: Used by ≤ 20% of respondents Usually or Always: Hand — Doyle Index; Knee — Walk time, Ascent time; Hip — Intercondylar distance, Intermalleolar distance, Walk time, Ascent time.

Arthritis Impact Measurement Scales (AIMS)[8], AIMS2[9], Rapid Assessment of Disease Activity in Rheumatology (RADAR)[10], Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)[11], Lequesne Indices of Clinical Severity[12], Dougados Functional Index[13], McGill Pain Questionnaire[14], Health Utilities Index[15], Nottingham Health Profile[16], Medical Outcome Survey Short Form-36[17], European Quality of Life Index[18], Index of Wellbeing[19], and Sickness Impact Profile[20], the reported frequency of usage was between 0 and 4%.

Data recording. Respondents differed in the procedures they usually followed for recording scores derived by the above instruments. The following methods were used: written notes in the patient's chart (34%), dictated notes in the patient's chart (31%), not recorded (21%), recorded on rough notes (6%), recorded on a flow sheet (4%), other (4%). When serially documenting the distribution of joint involvement in RA for longitudinal followup, a homunculus or mannikin was used Usually or Always by 63% of respondents (Always 36%, Usually 27%, Occasionally 21%, Never 16%).

Characteristics of a measure for use in adult outpatient

Table 4. Responses to the question: How often do you serially use the following assessment techniques for longitudinally monitoring the efficacy of antirheumatic drug therapy in your adult AS outpatient practice?

Assessment Technique	Never	Occasionally	Usually	Always
Schober test (or modification)*	5	14	34	47
Chest expansion*	4	18	35	43
Wall to occiput distance	10	22	34	34
Finger to floor distance	20	19	27	34
Sacroiliac joint tenderness	22	24	24	30
Finger to fibula distance	48	19	13	20
Chin to sternum distance	49	22	15	14

*Used by ≥ 70% of respondents Usually or Always.
NB: Used by ≤ 20% of respondents Usually or Always: Occiput to C7 distance, Tragus to wall distance, Dougados Articular Index.

Table 5. Responses to the question: How often do you serially use the following assessment techniques for longitudinally monitoring the efficacy of antirheumatic drug therapy in your adult FM outpatient practice?

Assessment Technique	Never	Occasionally	Usually	Always
Quality of sleep*	11	11	28	50
Fatigue*	13	13	29	45
No. tender points	17	15	29	39
Skinfold tenderness	47	25	19	9

*Used by ≥ 70% of respondents Usually or Always.
NB: Used by ≤ 20% of respondents Usually or Always: Reactive hyperemia, Dolorimeter scores.

practice. Respondents were asked to rate the importance of various characteristics, relevant to the use of measurement techniques in routine clinical practice, according to the following scale: "Very Important," "Quite Important," "Somewhat Important," "Quite Unimportant," "Not Important At All" (Table 6). Over 80% of respondents identified the same 6 characteristics of a measure for use in routine practice as being Very Important. The remaining respondents (with one exception) rated the same 6 characteristics as Quite Important (12–17%) or Somewhat Important (1–3%). The 6 characteristics were as follows: Simplicity, Quick Completion, Easy Scoring, Reliability, Validity, and Responsiveness. In general, use of a measure in prior clinical research studies or recommendation by the ACR for research purposes was regarded with higher levels of importance than recommendation for research use by the Food and Drug Administration (FDA), International League of Associations for Rheumatology (ILAR), World Health Organization (WHO), or European League of Associations for Rheumatology (EULAR).

Information recorded in the patient notes at outpatient visits. Change in arthritis status, response to treatment, current drug profile, and recent drug side effects were recorded by 100% of respondents, while change in overall health (93%) was almost always recorded. For drug monitoring, blood count (92%), biochemistry (89%), urinalysis (84%), and for inflammatory disease erythrocyte sedimentation rate (82%) were frequently recorded. Blood pressure (69%), pulse rate (38%), respiratory rate (11%), and weight (44%) were recorded less frequently.

Familiarity with basic measurement methods. Eighty-six percent of respondents thought Likert-type (i.e., Descriptive, Adjectival) scales were easy to use. A slightly smaller percentage (77%) felt that visual analog scales (VAS) were easy to use. Less than half the respondents (43%) felt comfortable using health status questionnaires, which required several items to be added together to give a final score. More respondents expressed adequate familiarity with the sphygmomanometric measurement of grip strength (94%) and the use of the goniometer (84%) than the dolorimeter (45%).

DISCUSSION

Quantitation of the clinical effects of interventions has become a standard procedure in clinical research. The methods used are based on valid, reliable, and responsive measurement techniques. The data produced are a required part of the licensing process for new antirheumatic drugs, similar procedures being used in the assessment of new orthopedic surgical techniques, the evaluation of physiotherapy modalities, and in health economics evaluations. Quantitative clinical measurement in routine clinical care

Table 6. Responses to the question: Consider measurements you use, or would like to use, in your clinical adult outpatient practice. How important would you generally rate the following characteristics?

Characteristic	Very Important	Quite Important	Somewhat Important	Quite Unimportant	Not Important At All
Previously used in clinical research studies	25	31	23	12	9
Recommended for research by ACR	16	22	30	17	15
Recommended for research by FDA	6	12	34	24	24
Recommended for research by ILAR	4	13	35	23	25
Recommended for research by WHO	3	12	34	27	24
Recommended for research by EULAR	2	15	35	22	26

offers several advantages: (1) it provides information regarding the severity of the patient's disease, and places the patient on the spectrum of disease; (2) it provides information to both physician and patient regarding the necessity to initiate, continue, modify, or terminate a particular therapy; (3) it provides information to disability insurers regarding the severity of disease and the outcome of treatment programs; (4) it provides information to litigators regarding the patient's health status and may provide some insight into attribution issues; and (5) it allows health care agencies to understand the clinical effect of their expenditures and, therefore, the appropriateness of ongoing payment for clinical interventions[21].

Response profile. Surveys have 2 potential limitations. First, if the response rate is low, the results may not be generalizable. However, in this survey, the response rate was extremely high (85%), and in as much as practice patterns may be a function of year of graduation[22], there was no significant difference between respondents and non-respondents. We believe, therefore, that our observations are generalizable to practice patterns in Canadian rheumatology. Second, surveys based on self-administered questionnaires provide information on what respondents say they do, and this may differ from what they actually do in practice. While this is an inherent weakness of survey techniques in general, a direct chart audit might not provide entirely accurate data with respect to the focus of this survey, since 21% of respondents indicated they did not record the information at all, while a further 10% made no formal record in the patient's chart. We feel that the postal survey methodology used provides reasonable insight into which outcome measures are currently used in routine care.

Rheumatoid arthritis. The measurement procedures most frequently used to longitudinally monitor RA outpatients were similar to those currently recommended by the ACR for Phase III clinical research studies (i.e., number of tender joints, number of swollen joints, pain, function, patient global assessment, and physician global assessment)[2]. Separate counts of tender and swollen joints (or, alternatively, involved or damaged joints) were preferred over graded joint counts, e.g., Ritchie Index[23]. However, despite being the major symptoms of RA, pain was routinely measured by less than 70% of respondents, and functional capacity was routinely assessed by less than 50% of respondents, the latter being more often assessed using the ACR Functional Classification[4] than the more responsive HAQ[5]. Comparative (same, better, worse) global scales were slightly more popular than Likert global scales, but both were distinctly more popular than global VAS. It is notable that the duration of morning stiffness was an extremely popular measure in clinical practice, yet it is not included in the ACR core set[2]. This may represent the fact that academics still debate the cause and nature of joint stiffness, whether patients can differentiate stiffness from pain, and whether it can be accurately quantitated given the difficulty of defining the point of first awakening (or getting out of bed) and the time of resolution (i.e., first noticeable improvement, or time when patient is as limber as they will be for the rest of the day). From a clinical care standpoint, however, the duration of morning stiffness is a very useful measure, since it varies as a function of disease activity and response to treatment. It is of note that performance based measures of physical function are not included in the ACR core set[2], and neither were they routinely used by the respondents to this survey.

The pattern of usage of different measures was similar for measuring the response to DMARD, NSAID, and corticosteroids. However, there was a trend toward the more frequent usage of outcome measures in assessing the response to DMARD therapy. This is not surprising, given the symptom modifying potential of all 3 classes of drugs, but also an expectation of substantial longterm benefits on joint structure and function from DMARD therapy. Finally, despite the availability of several well validated measures of quality of life, such measures were rarely used by respondents to the survey, and neither are they included in the ACR core set[2].

Osteoarthritis. Current clinical measurement procedures recommended by the Osteoarthritis Research Society (OARS), for hand, hip, and knee studies[3] include pain, physical function, and patient global assessment. In addition, physician global assessment is recommended and the measurement of quality of life highly recommended. The OARS guidelines[3], and the Outcome Measures in Arthritis Clinical Trials III recommendations[24] that preceded them, were published after the OMIRROP survey was conducted. In addition, because of space limitations, physical function measurement questions were not included in the generalized OA section, and global assessment questions were not included in the hip or knee sections.

In generalized OA, patient and physician global assessments were the only measures used Always or Usually by ≥70% of respondents, and comparative scales (same, better, worse) were slightly more popular than Likert scales. For pain measurement in generalized OA, Likert scales were used more often than VAS. With the exception of the ACR Functional Classification[4], health status questionnaires and health related quality of life questionnaires were rarely used. Simple counts of tender, swollen, damaged, or involved joints were most frequently employed, while the Doyle Index[25] (a graded count based on the Ritchie Index[23]) was rarely used. Grip strength and walk time were not usually measured in routine clinical care.

Despite being the principal symptom of hand, hip, and knee OA, pain was routinely measured by <70% of respondents. Nevertheless, measures of functional capacity were the most popular outcomes followed. Global assessments were surveyed only for hand OA and were used routinely by

>70% of respondents. Since global assessments were used routinely in generalized OA and hand OA, but the question was not asked for hip and knee OA, we speculate that the patient and physician global assessments may also be a popular method of monitoring patients with hip and knee OA. With 2 exceptions, all other types of measurement procedures were infrequently employed. The reasons for this were not surveyed but the measurement of stiffness is contentious, and joint counts are not particularly useful in regional conditions. Although the vast majority of respondents were familiar with the use of the modified sphygmomanometer and the goniometer, they did not use these instruments in longitudinal monitoring. In contrast, the assessment of joint swelling and crepitus are part of the routine examination of the musculoskeletal system, and it is not surprising that they form part of routine clinical monitoring.

Ankylosing spondylitis. There is currently no agreement on a core set of outcome measures for future AS clinical trials, although a working group recently has proposed a preliminary set[26]. The AS section of the OMIRROP survey questionnaire focused on examination based measures only. Since restricted lumbar spinal movement and chest expansion form part of the diagnostic criteria for AS[27], it is not surprising that they are also commonly used in outcome measurement. Many of the examination techniques used to assess AS patients need to be performed in a standardized fashion and both intraobserver and intrapatient variability taken into account[1]. Paradoxically, respondents to this survey were more likely to use the ACR Functional Classification[4] to rate patients with AS than to use the purpose-built Dougados Functional Index[13]. It should be noted that we did not specifically question rheumatologists regarding their use of the HAQ for the Spondyloarthropathies[28] or the Bath Ankylosing Spondylitis Functional Index (BASFI)[29].

Fibromyalgia. There is no international agreement on outcome measurement procedures for future Phase III FM studies[1]. The OMIRROP survey probed 6 outcome measures used in previous FM studies, but did not survey the use of pain or global measures. Given that nonrestorative sleep and chronic fatigue are key symptoms of FM[30], it is not surprising that they were the outcome measures most frequently employed by survey respondents. The number of tender points forms part of the diagnosis of FM[30], and almost 70% of respondents used them in longitudinal monitoring. The dolorimeter was rarely used, and this may be partly explained by the fact that 55% of respondents did not feel sufficiently familiar with its use. Neither the ACR Functional Classification[4] nor the HAQ[5] was developed specifically for use in patients with FM, and yet both were used more frequently than the purpose-built FM Impact Questionnaire[7].

It should be noted that quality of life measures were rarely used by respondents in monitoring patients with RA, OA, AS, or FM. Since there is still debate on the content and conceptual basis of health related quality of life questionnaires, concern that comorbidities may affect the score, and recognition that antirheumatic treatment may be successful but not alter the score, it is not surprising that practitioners have not yet adopted these measures for routine usage.

We conclude that most Canadian rheumatologists perform outcome assessment in their longitudinal followup of patients with RA, AS, OA, and FM. The usual method employed by the majority of respondents was to a large extent based on overall impressions rather than ratings derived from precise scores generated on a number of separate component measures. They agree on the qualities they require in an outcome assessment technique (quick completion, simplicity, easy scoring, validity, reliability, responsiveness) for use in routine outpatient care. The reason for the observed lack of standardization in the methods currently used is speculative. We suspect that it may be, in part, due to the following: (1) the lack, until recently, of international agreement on core sets of outcome measures; (2) a lack of familiarity with the format, administration, scoring, and interpretation of newer health status instruments; (3) the logistic constraints of performing some measurement procedures in a busy outpatient clinic setting; (4) the lack of any requirement to perform formalized serial measurements; (5) an absence of clinical research examining whether the availability of quantitative data makes a significant and favorable difference in clinical decision making, resource utilization, economic aspects of health care delivery, and, most importantly, in patient outcome; and (6) the lack of emphasis on formalized measurement in many rheumatology training programs. As a result, a rheumatologist may be more likely to measure the blood pressure, pulse rate, respiratory rate, and weight than to administer a generic or disease specific health status instrument or measure the patient's quality of life.

There is an opportunity to perform quantitative outcome measurement in rheumatology. To achieve this goal 3 requirements need to be met: (1) rheumatologists need to become familiar with the newer generic and disease specific self-administered health status questionnaires, (2) instrument developers need to address the measurement needs of rheumatologists and their patients in the routine outpatient clinical care setting, and (3) the benefits of practising quantitative clinical rheumatology need to be assessed from a number of standpoints (physician decision making, quality of care, resource utilization, health economics, and individual patient outcomes).

REFERENCES

1. Bellamy N. Musculoskeletal clinical metrology. Dordrecht: Kluwer Academic Publishers; 1993.
2. Felson DT, Anderson JJ, Boers M, et al. The American College of Rheumatology preliminary core set of disease activity measures for

rheumatoid arthritis clinical trials. Arthritis Rheum 1993;36:729-40.
3. Altman R, Brandt K, Hochberg M, et al. Design and conduct of clinical trials in patients with osteoarthritis: Recommendations from a task force of the Osteoarthritis Research Society. Osteoarthritis Cart 1996;4:217-43.
4. Hochberg MC, Chang RW, Dwosh I, Lindsey S, Pincus T, Wolfe F. The American College of Rheumatology 1991 revised criteria for the classification of global functional status in rheumatoid arthritis. Arthritis Rheum 1992;35:498-502.
5. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. Arthritis Rheum 1980;23:137-45.
6. Jette AM, Deniston OL. Interobserver reliability of a functional status assessment instrument. J Chron Dis 1978;31:573-80.
7. Burckhardt CS, Clark SR, Bennett RM. The Fibromyalgia Impact Questionnaire: Development and validation. J Rheumatol 1991;18:728-33.
8. Meenan RF, Gertman PM, Mason JH, Dunaif R. The Arthritis Impact Measurement Scales: Further investigations of a health status measure. Arthritis Rheum 1982;25:1048-53.
9. Meenan RF, Mason JH, Anderson JJ, Guccione AA, Kazis LE. AIMS2. The content and properties of a revised and expanded Arthritis Impact Measurement Scales Health Status Questionnaire. Arthritis Rheum 1992;35:1-10.
10. Mason JH, Anderson JJ, Meenan RF, Haralson KM, Lewis-Stevens D, Kaine JL. The Rapid Assessment of Disease Activity in Rheumatology (RADAR) Questionnaire. Validity and sensitivity to change of a patient self-report measure of joint count and clinical status. Arthritis Rheum 1992;35:156-62.
11. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt L. Validation study of WOMAC: A health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988;15:1833-40.
12. Lequesne M. Indices of severity and disease activity for osteoarthritis. Semin Arthritis Rheum 1991;20 Suppl 2:48-54.
13. Dougados M, Gueguen A, Nakache J-P, Nguyen M, Mery C, Amor B. Evaluation of a functional index and an articular index in ankylosing spondylitis. J Rheumatol 1988;15:302-7.
14. Melzack R. The McGill Pain Questionnaire: Major properties and scoring methods. Pain 1975;1:275-99.
15. Feeny D, Furlong W, Barr RD, Torrance GW, Rosenbaum P, Weitzman S. A comprehensive multiattribute system for classifying the health status of survivors of childhood cancer. J Clin Oncol 1992;10:923-8.
16. Hunt SM, McEwen J, McKenna SP. Measuring health status: a new tool for clinicians and epidemiologists. J Roy Coll Gen Pract 1985;35:185-8.
17. Ware JE, Sherbourne CD. The MOS 36-item Short-Form health survey (SF-36): I. Conceptual framework and item selection. Med Care 1992;30:473-83.
18. Hurst NP, Jobanputra P, Hunter M, Lambert M, Lochhead A, Brown H. Economic and Health Outcome Research Group. Validity of EuroQol — a generic health status instrument — in patients with rheumatoid arthritis. Br J Rheumatol 1994;33:655-62.
19. Kaplan RM, Bush JW, Berry CC. Health status: Types of validity and the index of well-being. Health Serv Res 1976;11:478-507.
20. Bergner M, Bobbitt RA, Pollard WE, Martin DP, Gilson BS. The Sickness Impact Profile: validation of a health status measure. Med Care 1976;14:57-67.
21. Bellamy N. Principles of outcome assessment. In: Klippel JH, Dieppe PA, eds. Rheumatology, 2nd ed. London: Mosby International 1998; vol. 1, section 3: 14-1–14.10.
22. Sackett DL, Haynes RB, Gibson ES. Clinical determinants of the decision to treat primary hypertension (abstract). Clin Res 1977;24:648A.
23. Ritchie DM, Boyle JA, McInnes JM, et al. Clinical studies with an articular index for the assessment of joint tenderness in patients with rheumatoid arthritis. Q J Med (New Series XXXVII) 1968; 147:393-406.
24. Bellamy N, Kirwan J, Boers M, et al. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. J Rheumatol 1997;24:799-802.
25. Doyle DV, Dieppe PA, Scott J, Huskisson EC. An articular index for the assessment of osteoarthritis. Ann Rheum Dis 1981;40:75-8.
26. van der Heijde DMFM, Bellamy N, Calin A, et al. Preliminary core sets for efficacy endpoints in ankylosing spondylitis. J Rheumatol 1997;24:2225-9.
27. van der Linden S, Valkenburg HA, Cats A. Evaluation of diagnostic criteria for ankylosing spondylitis: A proposal for modification of the New York criteria. Arthritis Rheum 1984;27:361-8.
28. Daltroy LH, Larson MG, Roberts WN, Liang MH. A modification of the Health Assessment Questionnaire for the spondyloarthropathies. J Rheumatol 1990;17:946-50.
29. Calin A, Garrett S, Whitelock H, et al. A new approach to defining functional ability in ankylosing spondylitis: The development of the Bath Ankylosing Spondylitis Functional Index (BASFI). J Rheumatol 1994;21:2281-5.
30. Wolfe F, Smythe HA, Yunus MB, et al. The American College of Rheumatology 1990 criteria for the classification of fibromyalgia: report of the multicenter criteria committee. Arthritis Rheum 1990;33:160-72.

A Survey of Outcome Measurement Procedures in Routine Rheumatology Outpatient Practice in Australia

NICHOLAS BELLAMY, KENNETH D. MUIRDEN, PETER M. BROOKS, DAVID BARRACLOUGH, MICHELLE M. TELLUS, and JANE CAMPBELL

ABSTRACT. Objective. To assess the extent to which quantitative clinical measurement is performed by rheumatologists in the longitudinal followup of patients with rheumatoid arthritis (RA), osteoarthritis (OA), ankylosing spondylitis (AS), and fibromyalgia (FM) in routine outpatient practice in Australia.

Methods. A cross sectional postal survey was conducted using an 18-item self-administered questionnaire sent to Australian Rheumatology Association (ARA) members.

Results. Rheumatologists (response rate = 76%, completion rate = 72%) were more likely to longitudinally follow patients with RA and AS than those with OA or FM. There was a high degree of variability in the methods used to monitor patients longitudinally. Many measures used in clinical research were used infrequently in routine clinical practice. In general, the major health status measures surveyed were not used in clinical monitoring. There was a high level of agreement (> 80%) that the characteristics required of an outcome measure for use in clinical practice should include simplicity, brevity, ease of scoring, reliability, validity, and sensitivity to change.

Conclusion. The majority of Australian rheumatologists perform outcome measurement during the longitudinal followup of their outpatients with RA, AS, OA, and FM. However, the process lacks standardization. High performance health status measures developed for clinical research have not been widely adopted in rheumatology practices. There is agreement on the characteristics required by Australian rheumatologists for measurement procedures used in routine clinical care. Quantitative measurement in clinical practice using standardized procedures is an attainable, but as yet, unrealized opportunity. (J Rheumatol 1999;26:1593–9)

Key Indexing Terms:
OUTCOME MEASUREMENT CLINICAL PRACTICE RHEUMATIC DISEASES

In recent years, there has been steady progress in the development of measurement techniques for clinical research purposes[1]. Accepted methodologies have been established for the development and validation of new measurement procedures. This evolution has resulted in the availability of a wide variety of outcome measurement alternatives for musculoskeletal clinical trials. One consequence of this development has been a lack of adequate standardization, different measures often being used in different studies[1]. Even with respect to primary outcomes, international agreement has been reached only recently on core sets of measures for future Phase III clinical trials in rheumatoid arthritis (RA)[2], and hip, knee and hand osteoarthritis (OA)[3] and ankylosing spondylitis (AS)[4].

While publications on measurement in routine clinical care are numerous[5-14], no surveys of practice habits of rheumatologists in conducting outcome measurement in routine clinical care have been published. Indeed, few techniques have been developed for specific application in the clinical practice setting[15,16]. We therefore surveyed monitoring practices used by Australian rheumatologists in longitudinal followup of patients with RA, OA (hip, hand, knee, and generalized), AS, and FM in routine clinical care. Our purpose was to describe current monitoring practices, determine required characteristics of instruments suitable for use in clinical practice, and gauge the extent to which several major health status instruments are currently being used in the clinical care setting. This survey was directed at monitoring patients and did not enquire whether rheumatologists acquired information for prognostic purposes.

MATERIALS AND METHODS

An 18 item (178 subcomponent) questionnaire was developed, pretested, revised, formatted, and then mailed in Australia to Australian rheumatologists. The survey was termed the Outcome Measurement In Rheumatology Routine Outpatient Practice (OMIRROP) Survey. The sample was ascertained from the 1995 Australian Rheumatology Association (ARA) Directory. Any ARA registrant identified who was not a clinical rheuma-

From the Department of Medicine, The University of Western Ontario, London, Canada, The University of Melbourne, Melbourne, The University of New South Wales, Sydney, and University of Queensland, Brisbane, Australia.

Supported in part by the Arthritis Foundation of Victoria and the University of Melbourne through the H&L Hecht Visiting Professorship.

N. Bellamy, MD, MSc, FRCP(Glas), FRCP(Edin), FRCPC, FACP, Professor of Rehabilitation Medicine, Department of Medicine, University of Queensland; K.D. Muirden, MD, BS, FRACP, Department of Medicine, University of Melbourne; D. Barraclough, MB, BS, FRACP; M.M. Tellus, MB, BS, FRACP, Royal Melbourne Hospital, University of Melbourne; P.M. Brooks, MD, FRACP, Executive Dean, Health Sciences, University of Queensland, Brisbane, Australia; J. Campbell, BA, Research Assistant, Department of Medicine, University of Western Ontario.

Address reprint requests to Prof. N. Bellamy, University of Queensland, Department of Medicine, Floor C, Clinical Sciences Building, Royal Brisbane Hospital, Herston, Queensland 4029, Australia.

Submitted February 20, 1998 revision accepted September 1, 1998.

tologist (e.g., an immunologist), or who was a pediatric rheumatologist, or not currently residing in Australia was excluded. A total of 197 eligible practicing rheumatologists was identified. Participants were questioned specifically regarding their measurement practices in the longitudinal followup (serial assessments over time) of their adult outpatients with RA, OA, AS, and FM. Because of the large number of outcome measures currently available, it was not possible to include all of them in the questionnaire. However, the majority were incorporated. For the purpose of the survey, an outpatient was defined as a nonhospitalized (i.e., ambulatory) patient seen in either private clinical practice or in the outpatient clinic of a health care facility. Second and third mailings of the OMIRROP questionnaire were made to nonrespondents at intervals of about one month, with a personal letter accompanying the third mailing to maximize the response rate.

Analysis was based mainly on descriptive statistics. To identify measurement procedures that were routinely used by a high percentage of respondents, we separately identified those outcome measures used "Always" or "Usually" by ≥ 70% of respondents. While this is an arbitrary division, nevertheless, it defines those measures that might be considered part of usual care.

RESULTS

Response data. Responses were obtained from 142 rheumatologists (response rate = 72%). The mean year of graduation from medical school of respondents was 1972 (range 1939–1986) (nonrespondent: mean = 1971, range 1949–1986, p = NS), and the mean year of starting practice in rheumatology of respondents was 1981 (range 1948–1995). The type of practice of respondents was as follows: Visiting Medical Officer (Non-Teaching Hospital) = 64%, Full Time Hospital = 24%, Visiting Medical Officer (Teaching Hospital) = 12%. The majority of respondents (93%) had experience participating in at least one prior clinical research project, in which they had been required to make or supervise clinical measurements on study subjects. Respondents were more likely to longitudinally follow patients with RA (99%) and AS (96%) than patients with knee OA (62%), hip OA (60%), generalized OA (58%), hand OA (49%), or FM (46%).

Basic measurement procedures. To assess the use of basic measurement procedures, participants were asked to respond to a number of questions in the following format: "How often do you serially use the following assessment techniques for longitudinally monitoring the efficacy of antirheumatic drug therapy in your adult (specify disease) outpatient practice?" Responses to this question were separately obtained with respect to RA (Table 1), generalized OA (Table 2), hip OA, knee OA and hand OA (Table 3), AS (Table 4), and FM (Table 5). For RA, participants were questioned also regarding monitoring practices in the separate situations of nonsteroidal therapy, disease modifying antirheumatic drug (DMARD) therapy, and steroid therapy (Table 1).

Usage patterns for RA, OA, AS, and FM varied for different outcome measures (Tables 1–5). Those used "Usually" or "Always" by ≥ 70% are identified by an asterisk and those used by ≤ 20% have been relegated to a footnote (Tables 1–5). Measures have been ranked according to the prevalence of "Always"/"Usually" usage (Tables 1, 2, 4, 5).

Health status instruments. No major health status instrument evaluated was used frequently in routine clinical practice. The reported frequency of usage for the American College of Rheumatology Functional Classification (FIQ)[1], Health Assessment Questionnaire (HAQ)[1], Functional

Table 1. Responses to the questions: (A) How often do you serially use the following assessment techniques for longitudinally monitoring the efficacy of antirheumatic drug therapy in your adult rheumatoid arthritis outpatient practice? (B) In what situation do you use the aforementioned assessment techniques in rheumatoid arthritis (please only answer for those techniques that you do use).

Assessement Technique	A				B		
	Nev	Occ	Usl	Alw	Monitoring NSAID Therapy	Monitoring DMARD Therapy	Monitoring Steroid Therapy
Duration of morning stiffness*†	2	4	41	54	70	87	77
Physician global assessment (same/better/worse)*†	5	4	28	62	73	85	78
Patient global assessment (same/better/worse)*†	5	5	31	58	69	83	76
Severity of morning stiffness	21	10	32	37	65	78	71
Physician global assessment (none/mild/moderate/severe)	23	13	23	41	62	74	65
No. of swollen joints	19	20	42	19	60	80	68
No. of involved joints	24	15	43	18	55	67	61
Patient global assessment (none/mild/moderate/severe)†	33	10	25	33	42	75	69
No. of tender joints	21	26	37	16	54	76	64
Pain scale (adjectival)	40	11	31	18	57	70	63
No. of damaged joints	31	27	34	8	38	59	52
Grip strength	38	30	20	12	43	59	52

Nev: Never; Occ: Occasionally; Usl: Usually; Alw: Always; NSAID: nonsteroidal antiinflammatory drug; DMARD: disease modifying antirheumatic drug. †Due to rounding percentage does not add to 100. *Used by ≥ 70% of the respondents Usually or Always.
NB: Used by ≤ 20% of respondents Usually or Always: Pain scale (VAS), Physician global assessment (VAS). Some other form of joint count (specify), Patient global assessment (VAS), Walk time, Ritchie articular index, ARA joint count, 28 joint count, Some other form of pain scale (specify).

Table 2. Responses to the question: How often do you serially use the following assessment techniques for longitudinally monitoring the efficacy of antirheumatic drug therapy in adult outpatients with generalized osteoarthritis?

Assessment Technique	Never	Occasionally	Usually	Always
Physician global assessment (same/better/worse)*	18	12	38	32
Patient global assessment (same/better/worse)	19	12	36	33
Physician global assessment (none/mild/moderate/severe)†	38	10	30	23
Pain scale (adjectival)†	45	8	32	16
Patient global assessment (none/mild/moderate/severe)	43	12	24	21
No. of involved joints†	41	17	30	11
No. of damaged joints†	48	13	29	11
No. of swollen joints	45	19	27	9
No. of tender joints†	50	20	24	7
Duration of morning stiffness	49	24	18	9
Severity of morning stiffness	58	16	20	6

*Used by ≥ 70% of respondents Usually or Always.
†Due to rounding does not add to 100.
NB: Used by ≤ 20% of respondents Usually or Always: Pain scale (VAS), Grip strength, Walk time, Physician global assessment (VAS), Patient global assessment (VAS), Some other form of joint count (specify), Some other form of pain scale (specify), ARA joint count, 28 joint count, Doyle articular index.

Table 3. Responses to the question: How often do you serially use the following assessment techniques for longitudinally monitoring the efficacy of antirheumatic drug therapy in your adult osteoarthritis outpatient practice in patients with (A) Hand OA, (B) Knee OA, and (C) Hip OA?

	(A) HAND OA				(B) KNEE OA				(C) HIP OA			
Assessment Technique	Nev	Occ	Usl	Alw	Nev	Occ	Usl	Alw	Nev	Occ	Usl	Alw
Functional capacity†*	21	10	46	24	15	12	48	26	16	6	50	28
Pain scale (VAS or adjectival)†	44	11	26	19	33	16	30	21	32	15	31	23
Duration of morning stiffness†	52	23	17	8	46	30	16	7	49	27	17	7
Severity of morning stiffness†	57	22	13	9	57	22	15	6	58	23	12	7
Patient global assessment*	18	11	37	34								
Physician global assessment†	20	12	36	33								
Number of involved joints	41	16	33	10								
Number of tender joints	40	20	32	8								
Number of swollen joints	39	22	29	10								
Grip strength	51	23	17	9								
Swelling (present/absent)*					13	4	47	36				
Crepitus (present/absent)†*					18	11	38	34				
Tenderness (present/absent)†					22	14	38	28				
Knee range of motion (by goniometry)					38	27	23	12				
Tenderness (graded)					57	18	16	9				
Hip range of motion (by goniometry)†									49	26	17	7

Nev: Never; Occ: Occasionally; Usl: Usually; Alw: Always. †Due to rounding percentage does not add to 100.
*Used by ≥ 70% of respondents Usually or Always.
NB: Used by ≤ 20% of respondents Usually or Always: Hand-Doyle Index; Knee-Walk time, Ascent time; Hip- -Intercondylar distance, Intermalleolar distance; Walk time, Ascent time.

Status Index[1] (FSI), Fibromyalgia Impact Questionnaire[1] (FIQ), Arthritis Impact Measurement Scales[1], Arthritis Impact Measurement Scales 2[1], Rapid Assessment of Disease Activity in Rheumatology[1], Western Ontario and McMaster Universities Osteoarthritis Index[1], Lequesne Indices of Clinical Severity[1], Dougados Functional Index[1], McGill Pain Questionnaire[1], Health Utilities Index[1], Nottingham Health Profile[1], Short Form-36 (SF-36)[1], European Quality of Life Index[1], Index of Wellbeing[1], and Sickness Impact Profile[1] was between 0% and 10%.

Data recording. Respondents differed in the procedures they usually followed for recording scores derived by the aforementioned measurement techniques. The following methods were used: written notes in the patient's chart (48%), dictated note in the patient's chart (16%), not recorded (26%), recorded on rough notes (7%), recorded on a flow

Table 4. Responses to the question: How often do you serially use the following assessment techniques for longitudinally monitoring the efficacy of antirheumatic drug therapy in your adult ankylosing spondylitis outpatient practice?

Assessment Technique	Never	Occasionally	Usually	Always
Chest expansion*	3	18	43	36
Schober test (or modification)*	9	20	40	31
Finger-to-floor distance	20	23	32	25
Sacroiliac joint tenderness	19	24	37	20
Wall–occiput distance	35	23	23	19

*Used by ≥ 70% of respondents Usually or Always.
NB: Used by ≤ 20% of respondents Usually or Always: Tragus-to-wall distance, Finger-to-fibula distance, Chin-to-sternum distance, Occiput-to-C7 distance, Dougados articular index.

Table 5. Responses to the question: How often do you serially use the following assessment techniques for longitudinally monitoring the efficacy of antirheumatic drug therapy in your adult fibromyalgia outpatient practice?

Assessment Technique	Never	Occasionally	Usually	Always
Quality of sleep*	18	8	42†	32
Fatigue†	21	14	43	23
No. of tender points	38	22	30	10

*Used by ≥ 70% of respondents Usually or Always.
†Due to rounding percentage does not add to 100.
NB: Used by ≤ 20% of respondents Usually or Always. Skinfold tenderness, Reactive hyperemia, Dolorimeter scores.

sheet (3%), other (2%). When serially recording the distribution of joint involvement in RA for longitudinal followup, a homunculus or mannikin was used "Usually" or "Always" by 14% of respondents (Always 5%, Usually 9%, Occasionally 34%, Never 52%).

Characteristics of a measure for use in adult outpatient practice. Respondents were asked to rate the importance of various characteristics, relevant to the use of measurement techniques in routine clinical practice, according to the following scale: "Very Important," "Quite Important," "Somewhat Important," "Quite Unimportant," "Not Important At All" (Table 6). Over 80% of respondents identified the same 6 characteristics of a measure for use in routine practice as being "Very Important." The remaining respondents (with one exception) rated the same 6 characteristics as "Quite Important" (11–17%), or "Somewhat Important" (1–3%). The 6 characteristics were as follows: Simplicity, Quick completion, Easy scoring, Reliability, Validity, and Responsiveness. In general, use of a measure in prior clinical research studies or recommendation by the American College of Rheumatology (ACR) for research purposes was regarded with higher levels of importance than recommendation for research use by the Food and Drug Administration, International League of Associations for Rheumatology, World Health Organization, or European League of Associations for Rheumatology.

Information normally recorded in the patient's notes at most outpatient visits. Response to treatment, current drug profile, recent drug side effects, change in arthritis status, and change in overall health were almost always recorded (92–99%). For drug monitoring, blood count, biochemistry, urinalysis, and for inflammatory disease erythrocyte sedimentation rate were also frequently recorded (85–96%). Blood pressure (48%), body weight (43%), pulse rate (22%), and respiratory rate (5%) were recorded less frequently.

Familiarity with basic measurement methods. Eighty-six percent of respondents were of the opinion that Likert-type (i.e., Descriptive, Adjectival) scales were easy to use. A slightly smaller percentage (75%) felt that visual analog scales (VAS) were easy to use. Less than half the respondents (47%) felt comfortable using health status questionnaires, which required several items to be added together to give a final score. More respondents expressed adequate familiarity with the use of the goniometer (85%) and the sphygmomanometric measurement of grip strength (83%) than with the use of the dolorimeter (26%).

Table 6. Responses to the question: Consider measurements that you use, or would like to use, in your clinical adult outpatient practice. How important would you generally rate the following characteristics?

Characteristic	Very Important	Quite Important	Somewhat Important	Quite Unimportant	Not Important At All
Previously used in clinical research studies	22	28	35	7	8
Recommended for research by ACR*	9	17	38	22	15
Recommended for research by FDA	5	11	36	28	20
Recommended for research by ILAR	3	17	40	23	17
Recommended for research by WHO*	3	12	40	24	20
Recommended for research by EULAR*	3	11	44	22	19

*Due to rounding percentage does not add to 100.

DISCUSSION

Quantitation of the clinical effects of interventions has become a standard procedure in clinical research. The methods used are based on valid, reliable, and responsive measurement techniques. The data produced are a required part of the licensing process for new antirheumatic drugs, similar procedures being used in the assessment of new orthopedic surgical techniques, the evaluation of physiotherapy modalities, and in health economics evaluations. In contrast, quantitative clinical measurement in routine clinical care offers several advantages: (a) it provides information regarding the severity of the patient's disease, places the patient on the spectrum of disease, and may be of prognostic importance[9,17-19]; (b) it provides information to both physician and patient regarding the necessity to initiate, continue, modify, or terminate a particular therapy; (c) it provides information to disability insurers regarding the severity of disease and the outcome of treatment programs; (d) it provides information to litigators regarding the patient's health status and may provide some insight into attribution issues; and (e) it allows health care agencies to understand the clinical effect of their expenditures and, therefore, the appropriateness of ongoing payment for clinical interventions[20].

Response profile. Surveys have 2 potential limitations. First, if the response rate is low, the results may not be generalizable. However, in this survey, the response rate was high (i.e., 72%), and in as much as practice patterns may be a function of year of graduation[21], there was no significant difference between respondents and nonrespondents. We believe, therefore, that our observations are generalizable to rheumatology practice patterns in Australia. Second, surveys based on self-administered questionnaires provide information on what respondents say they do, and this may differ from what they actually do in practice. While this is an inherent weakness of survey techniques in general, a direct chart audit might not provide entirely accurate data with respect to the focus of this survey, since 26% of respondents indicated that they did not record the information at all, while a further 9% made no formal record in the patient's chart. We feel that the postal survey methodology used provides reasonable insight into which outcome measures are used currently in routine care.

General observations. In general, comparative assessments (same, better, worse) were slightly preferred over quantitative measures. VAS were less popular than Likert scales for assessing health status. Standard health status measures were rarely used. Since there is still debate as to the content and conceptual basis of health related quality of life questionnaires[22], concern that comorbidities may influence the score[23], and recognition that antirheumatic treatment may be successful but not alter the score[24], it is not surprising that practitioners have not yet adopted these measures for routine use. Simple binary joint counts were more frequently used than graded joint counts. Tests of physical performance were rarely used.

Rheumatoid arthritis. Current ACR outcome measurement guidelines for Phase III clinical research studies recommend the assessment of number of tender joints, number of swollen joints, pain, function, patient global assessment, and physician global assessment[2]. While separate counts of tender and swollen joints were preferred over graded joint counts, in general, specific joint counting procedures were employed by < 70% of respondents. Furthermore, despite being the major symptom of RA, pain scales were routinely employed by less than 70%, and functional capacity assessments by less than 11%, of respondents, the majority of those using the HAQ (10%). It is of note that the duration of morning stiffness was an extremely popular measure in clinical practice, but yet is not included in the ACR core set (cf, physician and patient global assessments)[2]. In contrast, performance based measures are not included in the ACR core set[2], and were not routinely used by survey respondents.

There was a trend towards the more frequent usage of outcome measures in assessing the response to DMARD therapy. This is not surprising given the symptom-modifying potential of all 3 classes of drugs but an added expectation of substantial longterm benefits on joint structure and function from DMARD therapy.

Osteoarthritis. Current clinical measurement procedures, recommended by the Osteoarthritis Research Society (OARS), for hand, hip, and knee studies[3] include pain, physical function, and patient global assessment. In addition, physician global assessment is recommended and the measurement of quality of life highly recommended. The OARS guidelines[3] and the OMERACT III recommendations[25] that preceded them were published after the OMIRROP survey was conducted. In addition, because of space limitations, physical function measurement questions were not included in the generalized OA section, and global assessment questions were not included in the hip or knee sections.

In generalized OA, the physician global assessment predominated and was the only measure used "Always" or "Usually" by ≥ 70% of respondents.

Despite being the principal symptom of hand, hip, and knee OA, pain was routinely measured by < 70% of respondents. Nevertheless, measurements of functional capacity were popular outcome measures. Global assessments were surveyed only for hand OA and were used routinely by about 70% of respondents. Since global assessments were used routinely in generalized OA and hand OA, but the question was not asked for hip OA and knee OA, we speculate that the patient and physician global assessments may also be a popular method of monitoring hip and knee OA patients. Although the vast majority of respondents were familiar with the use of the modified sphygmomanometer

and the goniometer, they did not use these instruments in longitudinal monitoring. Quantitating joint swelling is notoriously difficult, while crepitus is not usually considered quantifiable and not expected to change with treatment.

Ankylosing spondylitis. A degree of agreement has been reached on a core set of outcome measures for future AS clinical trials. The AS section of the OMIRROP survey questionnaire focused on examination-based measures only. Since restricted lumbar spinal movement and chest expansion form part of the diagnostic criteria for AS, it is not surprising that they are commonly used also in outcome measurement. It is a paradox that respondents to this survey were more likely to use the FSI, HAQ, or the ACR Functional Classification in AS patients than the purpose-built Dougados Functional Index. It should be noted that we did not specifically question rheumatologists regarding their use of the HAQ-S[25] or the Bath Ankylosing Spondylitis Functional Index[26].

Fibromyalgia. There is currently no international agreement on outcome measurement procedures for future Phase III FM studies[1]. The OMIRROP survey probed 6 outcome measures used in previous FM studies, but did not survey the use of pain or global measures. Given that non-restorative sleep and chronic fatigue are key symptoms of FM[27], it is not surprising that they were the most frequently employed outcome measures by survey respondents. The number of tender points forms part of the diagnosis of FM[27] but only 40% of respondents used them in longitudinal monitoring. The dolorimeter was rarely used, and may be partly explained by the fact that 74% of respondents did not feel sufficiently familiar with its use. While the FSI was not developed specifically for use in FM patients, it was used more frequently than the purpose-built FIQ.

We conclude from this survey that most Australian rheumatologists perform outcome assessment in their longitudinal followup of their patients with RA, AS, OA, and FM. The usual methods employed by the majority of respondents were to a large extent based on overall impressions rather than ratings derived from precise scores generated on a number of separate component measures. They agree on the qualities required of an outcome assessment technique for use in routine outpatient care. The observed lack of standardization in the methods currently used is speculative. We suspect that it may be in part due to the following: (a) lack, until recently, of international agreement on core sets of outcome measures; (b) lack of familiarity with the format, administration, scoring, and interpretation of newer health status instruments; (c) the logistic constraints of performing some measurement procedures in a busy outpatient clinic setting; (d) the lack of any requirement to perform formalized serial measurements; (e) an absence of clinical research examining whether the availability of quantitative data has significant and favorable consequences in clinical decision making, resource utilization, economic aspects of health care delivery and, most importantly, on patient outcome; and (f) the lack of emphasis on formalized measurement in many rheumatology training programs. As a result, a rheumatologist may be more likely to measure the blood pressure, weight, and pulse rate, than to administer a generic or disease-specific health status instrument or measure the patient's quality of life.

There is an opportunity to perform quantitative outcome measurement in rheumatology. To achieve this goal 4 requirements need to be met: (a) rheumatologists need to become familiar with the newer generic and disease-specific self-administered health status questionnaires; (b) instrument developers need to address the measurement needs of rheumatologists and their patients in the routine outpatient clinical care setting; (c) the consequences of practicing quantitative clinical rheumatology need to be assessed from a number of standpoints (physician decision making, quality of care, resource utilization, health economics, monitoring and prognosticating individual patient outcomes); and (d) patient management systems need to be developed that are user friendly, affordable, and capable of performing multiple functions (data management, booking, billing, etc.) that will reduce workload and enhance decision making for practicing rheumatologists and improve the precision and efficiency of clinical practice. Perhaps in practical terms this last element will be the critical step in the successful implementation of quantitative rheumatology. Finally, it is not that we need to replace existing outcome measures but it is necessary to address not only the clinimetric issues of validity and responsiveness but also the third element of the OMERACT Filter, that of feasibility[28]. This can only be achieved by bridging the gap between the theoretical and the practical and understanding the reality of performing quantitative measurement in clinical practice.

REFERENCES

1. Bellamy N. Musculoskeletal clinical metrology. Dordrecht: Kluwer Academic Publishers; 1993.
2. Felson DT, Anderson JJ, Boers M, et al: The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. Arthritis Rheum 1993;36:729-40.
3. Altman R, Brandt K, Hochberg M, et al. Design and conduct of clinical trials in patients with osteoarthritis: recommendations from a task force of the Osteoarthritis Research Society. Osteoarthritis Cart 1996;4:217-43.
4. van der Heijde D, Bellamy N, Calin A, Dougados M, Khan A, van der Linden S and the ASAS Working Group. Preliminary core sets for efficacy endpoints in ankylosing spondylitis. J Rheumatol 1997;24:2225-9.
5. Brooks RH, Callahan LF, Pincus T. Use of self-report activities of daily living questionnaires in osteoarthritis. Arthritis Care Res 1988;1:23-32.
6. Callahan LF, Brooks RH, Summey JA, Pincus T. Quantitative pain assessment for routine care of rheumatoid arthritis patients, using a pain scale based on activities of daily living and a visual analogue pain scale. Arthritis Rheum 1987;30:630-6.
7. Hawley DJ, Wolfe F. Sensitivity to change of the Health

Assessment Questionnaire (HAQ) and other clinical and health status measures in rheumatoid arthritis: results of short term clinical trials and observational studies versus long term observational studies. Arthritis Care Res 1992;5:130-6.
8. Pincus T. Why should rheumatologists collect patient self-report questionnaires in routine rheumatologic care? Rheum Dis Clin North Am 1995;21:271-317.
9. Pincus T. Consecutive patient questionnaire databases: a complementary alternative. J NIH Res 1997;9:32-9.
10. Pincus T, Callahan LF. Quantitative measures to assess, monitor and predict morbidity and mortality in rheumatoid arthritis. Baillieres Clin Rheum 1992;6:161-91.
11. Pincus T, Wolfe F. Patient self-report questionnaires as integral to clinical care. Health and psychological instruments, behavioral measurement database services. Behav Meas Ltr 1997;5:3-7.
12. Wolfe F, Kleinheksel SM, Cathey MA, Hawley DJ, Spitz P, Fries JF. The clinical value of the Stanford Health Assessment Questionnaire Functional Disability Index in patients with rheumatoid arthritis. J Rheumatol 1988;15:1480-8.
13. Wolfe F, Pincus T. Standard self-report questionnaires in routine clinical and research practice — an opportunity for patients and rheumatologists. J Rheumatol 1991;18:643-6.
14. Wolfe F, Pincus T. Data collection in the clinic. Rheum Dis Clin North Am 1995;21:321-58.
15. Wolfe F. A brief health status instrument: CLINHAQ [abstract]. Arthritis Rheum 1989;32:S99.
16. Pincus T, Callahan LF, Brooks RH, Fuchs HA, Olsen NJ, Kaye JJ. Self-report questionnaire scores in rheumatoid arthritis compared with traditional physical, radiographic, and laboratory measures. Ann Inten Med 1989;110:259-66.
17. Pincus T, Callahan LF. Taking mortality in rheumatoid arthritis seriously-predictive markers, socioeconomic status and comorbidity. J Rheumatol 1986;13:841-5.
18. Pincus T, Callahan LF. Rheumatology function tests: grip strength, walking time, button test and questionnaires document and predict longterm morbidity and mortality in rheumatoid arthritis. J Rheumatol 1992;19:1051-7.
19. Wolfe F, Cathey MA. The assessment and prediction of functional disability in rheumatoid arthritis. J Rheumatol 1991;18:1298-306.
20. Bellamy N. Principles of outcome assessment. In: Klippel JH, Dieppe PA, editors. Rheumatology. 2nd ed. London: Mosby.Vol.1.Section 3:14.1-14.10.
21. Sackett DL, Haynes RB, Gibson ES. Clinical determinants of the decision to treat primary hypertension. Clin Res 1977;24:648A.
22. Hunt S. Defining quality of life: the practical importance of conceptual clarity — technical, ethical, and interpretive issues. Medical Outcomes Trust Monitor 1997; April Vol 2, Issue 2. Boston.
23. Hawker G, Melfi C, Paul J, Green R, Bombardier C. Comparison of a generic (SF-36) and a disease specific (WOMAC) instrument in the measurement of outcomes after knee replacement surgery. J Rheumatol 1995;22:1193-6.
24. Bellamy N, Kirwan J, Boers M, et al. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. J Rheumatol 1997;24:799-802.
25. Daltroy LH, Larson MG, Roberts WN, Liang MH. A modification of the Health Assessment Questionnaire for the spondyloarthropathies. J Rheumatol 1990;17:946-50.
26. Calin A, Garrett S, Whitelock H, et al. A new approach to defining functional ability in ankylosing spondylitis: the development of the Bath Ankylosing Spondylitis Functional Index (BASFI). J Rheumatol 1994;21:2281-5.
27. Wolfe F, Smythe HA, Yunus MB, et al. The American College of Rheumatology 1990 criteria for the classification of fibromyalgia: report of the multicenter criteria committee. Arthritis Rheum 1990;33:160-72.
28. Boers M, Brooks P, Strand V, Tugwell P. The OMERACT Filter for outcome measures in rheumatology. J Rheumatol 1998;25:198-9.

EXTENDED REPORT

Using patients' and rheumatologists' opinions to specify a short form of the WOMAC function subscale

F Tubach, G Baron, B Falissard, I Logeart, M Dougados, N Bellamy, P Ravaud

Ann Rheum Dis 2005;**64**:75–79. doi: 10.1136/ard.2003.019539

See end of article for authors' affiliations

Correspondence to: Dr Florence Tubach, Département d'Epidémiologie, Biostatistique et Recherche Clinique, INSERM E0357, Hôpital Bichat, 46 rue Henri Huchard, 75018 Paris, France; florence.tubach@bch.ap-hop-paris.fr

Accepted 26 March 2004

Background: The WOMAC (Western Ontario and McMaster Universities) function subscale is widely used in clinical trials of hip and knee osteoarthritis. Reducing the number of items of the subscale would enhance efficiency and compliance, particularly for use in clinical practice applications.

Objective: To develop a short form of the WOMAC function subscale based on patients' and experts' opinions (WOMAC function short form).

Methods: WOMAC function subscale data (Likert version) were obtained from 1218 outpatients with painful hip or knee osteoarthritis. These patients and their rheumatologists selected the five items that they considered most in need of improvement. The rheumatologists were asked to select the five items for which patients in general are the most impaired. Items that were least important to patients and experts, those with a high proportion of missing data, and those with a response distribution showing a floor or ceiling response were excluded, along with one of a pair of items with a correlation coefficient >0.75.

Results: The WOMAC function short form included items 1, 2, 3, 6, 7, 8, 9, and 15 of the long form. The short form did not differ substantially from the long form in responsiveness (standardised response mean of 0.84 *v* 0.80).

Conclusions: A short form of the WOMAC function subscale was developed according to the views of patients and rheumatologists, based on the responses of 1218 patients and 399 rheumatologists. The clinical relevance and applicability of this WOMAC function subscale short form require further evaluation.

One of the major uses of health measurement scales is detecting health status change over time, either in the context of clinical trials or epidemiological studies or as a strategy for monitoring the outcomes and making decisions about the care of individual patients in daily clinical practice. In all situations, a priority may be efficiency, achieved by the shortest possible questionnaire.[1] To date, methods of shortening questionnaires have focused on approaches that maintain the greatest internal consistency.[2] However, in the context of health measurement scales targeting a relatively heterogeneous disorder, it may be advantageous to sacrifice internal consistency for content validity.[3]

The Western Ontario and McMaster Universities (WOMAC) osteoarthritis index is a valid, reliable, and responsive measure in hip and knee osteoarthritis.[4 5] This index is self administered and involves 17 items addressing the degree of difficulty in accomplishing 17 activities of daily life. While the mean importance score of the 17 items is similar at a group level, there is interindividual variability in the importance attached by individual patients to particular items.[4 5] The WOMAC function subscale is short, and can be completed quickly. Nevertheless, an even shorter version would further enhance its applicability in epidemiological studies and for use in routine clinical practice.[2]

Our aim in this study was to specify a short form of the WOMAC function subscale dedicated to all patients with hip or knee osteoarthritis, by preserving the most important items for patients and rheumatologists (WOMAC function subscale short form).

METHODS

Study population

We conducted a prospective cohort study of four weeks' duration, involving 1362 outpatients with hip or knee osteoarthritis as defined by the American College of Rheumatology,[6 7] and 399 private rheumatologists in France. Each rheumatologist was required to include four patients, three with knee osteoarthritis and one with hip osteoarthritis. To be included in the study, patients had to experience pain from the osteoarthritis (≥30 mm on a visual analogue scale (VAS) ranging from 0 to 100 mm) and to require treatment with a non-steroidal anti-inflammatory drug (NSAID). Inclusion could begin with the onset of treatment or with a switch from one NSAID to another. Patients were excluded if they had a prosthesis on the assessed joint or if they had been treated with intra-articular injection in the four weeks before the study began. All patients initially visited the rheumatologist in charge of their case, and an NSAID was prescribed (the drug and its dosage were chosen by the physician). A final visit to the same rheumatologist was scheduled four weeks later.

Measurements

Patients and rheumatologists assessed the patient's status with respect to the osteoarthritis at the baseline visit and at week 4. Patients completed the French Canadian version of the WOMAC physical function subscale[8] (17 items, five point Likert scale version, total score varying between 0 and 68; high scores indicate a high degree of functional impairment).

Patients were also asked to select the five items of the WOMAC function subscale that they considered most in need of improvement.

The rheumatologists were asked on one occasion to select the five items on the subscale which they consider result in the greatest impairment in patients with knee and hip osteoarthritis (not the specific patients they had included in the study).

Abbreviations: ICC, intraclass correlation coefficient; SRM, standardised response mean; WOMAC, Western Ontario and McMaster Universities osteoarthritis index

Table 1 Baseline characteristics of the patients

	Full WOMAC function subscale* (n = 1218)	Incomplete WOMAC function subscale† (n = 144)
Age (years)	66.9 (10.4)	69.7 (10.8)
Female sex	854 (70.1%)	59 (41.0%)
Body mass index (kg/m²)	27.6 (4.7)	27.8 (5.0)
Disease duration (years)	4.5 (5.6)	4.7 (5.9)
Kellgren and Lawrence grade		
II	246 (20.2%)	19 (13.2%)
III	530 (43.6%)	62 (43.1%)
IV	440 (36.2%)	63 (43.8%)
NSAID intake during the past 4 weeks	355 (29.2%)	38 (26.4%)
Analgesics‡ during the past 4 weeks	699 (57.5%)	87 (60.4%)
Symptomatic slow acting drug intake§ during the past 4 weeks	437 (36.0%)	38 (26.4%)
Pain (0–100 VAS)		
Mean (SD)	57.9 (17.0)	57.6 (17.4)
Week 0 to week 4 (SD)	−23.3 (22.1)	−21.4 (22.4)
Global assessment (0–100 VAS)		
Mean (SD)	58.6 (19.0)	59.3 (20.6)
Week 0 to week 4 (SD)	−23.1 (24.1)	−21.0 (26.6)
WOMAC function score (0–68)		
Mean (SD)	29.7 (11.4)	-
Week 0 to week 4 (SD)	−7.8 (9.7)	-

Values are mean (SD) or n (%).
*Patients without missing data for the WOMAC function subscale at the baseline visit; thus those who were involved in the derivation process.
†Patients with missing data for the WOMAC function subscale or who did not complete the questions at all at baseline.
‡Other than NSAID.
§Chondroitin sulphate, diacerhein, or avocado/soybean unsaponifiables.
NSAID, non-steroidal anti-inflammatory drugs; VAS, visual analogue scale; WOMAC, Western Ontario and McMaster Universities osteoarthritis index.

To assess the test–retest reliability of the resulting WOMAC function subscale short form, a subsample of 86 patients was asked to complete the full WOMAC function subscale again, 48 hours after the baseline visit. These patients had begun taking NSAIDs 48 hours after the baseline visit (that is, after completing the WOMAC function subscale a second time).

Statistical analysis

First, we computed descriptive statistics on clinical and demographic variables. Then we used a four step procedure to eliminate items as follows:

- Step 1. We ranked the 17 items of the complete WOMAC function subscale from highest to lowest importance according to the patients' and rheumatologists' opinions, excluding the five items that were least important for both patients and rheumatologists. The whole sample was then divided into tertiles of the WOMAC function subscale score to investigate the potential impact of the level of functional impairment on the patients' ranking.
- Step 2. We ranked the 17 items by the proportion of missing data per item. Items with a high proportion of missing data were excluded.
- Step 3. Items whose distribution of answers showed a floor or ceiling response were excluded. This response is present when most of the answers are clustered in only a few response options at one extreme—that is, when most of the subjects attest to having no difficulty (floor response) or extreme difficulty (ceiling response) in the activity. For floor response items, it is impossible to detect improvement, while for ceiling response items, it is not possible to distinguish among various grades of difficulty, as most of the subjects answer the same way.
- Step 4. We tested for potentially redundant items. Inter-item correlation coefficients were computed. When the correlation coefficient was greater than 0.75, the least important item of the pair in the patients' ranking was excluded.

Responsiveness was assessed by use of the standardised response mean (SRM) for the complete WOMAC function subscale and the WOMAC function short form. SRM is the mean change in score between the baseline and the final visit divided by the standard deviation of the change in score. Test–retest reliability was assessed using the intraclass correlation coefficient (ICC). Construct validity of the WOMAC function short form was assessed using the correlation between scores of the long and short forms, as recommended when the original scale cannot be considered a gold standard (that is, the reference measurement instrument).[2] Internal consistency was assessed using Cronbach's α.[9]

Statistical analyses involved use of the SAS Release 8.2 statistical software package.

RESULTS

In all, 1362 patients were enrolled in the study: 1019 (75%) with knee osteoarthritis and 343 (25%) with hip osteoarthritis. At the baseline visit, 1218 patients (89%) completed the full WOMAC function subscale without any missing data. The derivation process is based on these 1218 patients, described in table 1.

Ranking of the 17 items of the complete WOMAC function subscale

Patients and rheumatologists were consistent in ranking the importance of items (table 2). The four most important items for rheumatologists were among the five most important items for patients. The five least important items for rheumatologists were among the six least important items for patients. The ranking of item importance was similar

between patients with hip osteoarthritis and those with knee osteoarthritis (data not shown), except for "descending stairs" (ranked sixth and first, respectively), and "putting on socks/stockings" (first and 12th, respectively). As these items are relatively specific to the location of the osteoarthritis (hip or knee), this discrepancy was expected. The ranking of the items' importance was similar between men and women (data not shown), except for "going shopping" (ranked 10th and fourth, respectively) and "performing light domestic duties" (ranked 13th and seventh, respectively).

The five least important items for both patients and experts were "lying in bed," "bending to the floor," "rising from bed," "sitting," "taking off socks/stockings," and "standing."

Results of dividing the whole sample into tertiles of the WOMAC function subscale score showed exactly the same items being selected by patients in the three subgroups.

Ranking of the 17 items by the proportion of missing data

Three items generated notably more missing data than the others. These items may have been interpreted too literally and considered not to be relevant—for example, domestic duties may have been interpreted only as cleaning the house and therefore probably of more concern to women, while respondents answering the getting in/out of the bath question may not have appreciated that this question can alternatively be considered relevant to getting in/out of the shower.

The items excluded were "performing heavy domestic duties," "performing light domestic duties," and "getting in/out of the bath."

Items for which the distribution of answers showed a floor or ceiling response

Almost all the items of the complete WOMAC function subscale had a good distribution of answers among response modes. However, two had a saturation point in one or two response modes: for "bending to the floor" and "lying in bed," 74% and 75% of the answers, respectively, were "no difficulty" or "slight difficulty."

The items excluded were "bending to the floor" and "lying in bed." Both items had already been excluded in a previous step.

Inter-item correlation

Pairs of highly correlated items ($r>0.75$) were "putting on socks/stockings" with "taking off socks/stockings" ($r = 0.85$) and "performing light domestic duties" with "performing heavy domestic duties" ($r = 0.78$).

The items excluded were "taking off socks/stockings," and "performing heavy domestic duties." Both items had been excluded in a previous step.

Summary of the reduction procedure

The eight items of the WOMAC function subscale short form derived by the above mentioned methods are shown in the appendix. These items were the eight most important in the patients' opinion.

When summarising the different steps in the reduction procedure, it can be seen that six of the nine excluded items were excluded in at least two steps (two steps for four of the items and three for two of the items).

Performance characteristics

The WOMAC function subscale short form did not differ substantially from the complete WOMAC function subscale either in responsiveness (SRM = 0.84 (n = 1169) and 0.80 (n = 1048), respectively) or in test–retest reliability (ICC = 0.75 (0.65 to 0.83) and 0.79 (0.69 to 0.87), respectively).

Construct validity of the WOMAC function subscale short form was excellent ($r = 0.95$ between the long and short forms). Internal consistency was good in the WOMAC function subscale short form and the complete WOMAC function subscale ($\alpha = 0.84$ and $\alpha = 0.93$, respectively).

DISCUSSION

Using patients' and rheumatologists' opinions in France, and based on the Likert version of the French Canadian WOMAC function subscale, we have specified a short form of this subscale for patients with hip or knee osteoarthritis (including a broad spectrum of disease severity). To address recent recommendations for shortening composite measurement scales,[2] we have ensured that the original scale was valid, relevant in the context of hip and knee osteoarthritis, and had satisfactory measurement properties.[4 5]

The WOMAC function subscale short form contains only eight items. It was derived by preserving face validity

Table 2 The 17 items of the complete WOMAC function subscale ranked in importance by patients and rheumatologists

Item	Patients' opinions* (n = 1349)	Rheumatologists' opinions† (n = 497)	Patients' ranking‡	Rheumatologists' ranking§	Proportion of missing data (%)
Descending stairs	64.10	73.85	1	1	0.22
Ascending stairs	62.65	47.90	2	5	0.29
Walking on the flat	50.20	28.15	3	8	0.15
Getting in/out of a car	41.45	51.90	4	4	0.22
Rising from sitting	39.90	66.00	5	3	0.37
Going shopping	34.10	19.30	6	10	0.95
Getting on/off the toilet	33.65	28.95	7	7	0.37
Putting on socks/stockings	30.85	66.80	8	2	0.44
Getting in/out of the bath	30.00	45.65	9	6	4.26
Performing light domestic duties	26.65	10.25	10	12	4.77
Performing heavy domestic duties	26.30	23.75	11	9	5.29
Standing	25.70	9.05	12	13	0.22
Taking off socks/stockings	11.95	17.10	13	11	0.59
Sitting	8.40	3.20	14	15	0.37
Rising from bed	7.05	5.65	15	14	0.37
Bending to floor	4.60	2.00	16	16	0.22
Lying in bed	2.40	0.40	17	17	0.22

Each patient selected the five items they considered the most important to be improved by treatment, each rheumatologist selected the five items for which their patients are generally most impaired. Excluded items are in italics.
*Percentage of patients who considered this item as one of the five most important.
†Percentage of rheumatologists who considered this item as one of the five most important.
‡Rank of the item based on the percentage of patients who considered this item as one of the five most important.
§Rank of the item based on the percentage of rheumatologists who considered this item as one of the five most important.

(patients' and rheumatologists' opinions) and quality of the items (few missing data, no redundancy, good distribution of the answers across response modes). Preserving face validity is important because it increases the acceptance of the instrument by those who will ultimately use it and thus decreases the amount of missing data.[3] This short form has good responsiveness, good test–retest reliability, and good construct validity for this sample, but these parameters should be validated in an independent sample of subjects from the target population.[10] Our reduction procedure involved deleting items that were highly correlated, and thus a lower internal consistency was expected for the short form than for the complete subscale (an internal Cronbach's $\alpha = 1$ indicates redundancy).

As the WOMAC subscale is dedicated to patients with hip or knee osteoarthritis, our sample reflects this target population well. The proportion of patients with hip and knee osteoarthritis (three quarters knee and one quarter hip) is close to the distribution in the community.[11] As shown in table 1, our sample, is similar to samples included in trials on osteoarthritis treatment and represents a large spectrum of disease severity. Inclusion criteria, especially the requirement for NSAID treatment, were the same as those in the validation study of the WOMAC scale by Bellamy and associates.[4] In our sample, the same items were selected by patients across the range of osteoarthritis severity: the results of dividing the sample into tertiles of the WOMAC function subscale score showed that the five least important items to patients (those to be excluded) were exactly the same in the three tertiles.

It has been assumed that items for assessing knee osteoarthritis may be somewhat different from those required for hip disease. In fact, when we evaluated the ranking of the 17 items of the complete WOMAC function subscale according to their importance to patients with hip or knee disease, the five least important items (those to be excluded) were the same for patients with both types of osteoarthritis.

According to previous recommendations, when the original scale cannot be considered a gold standard (the reference measurement instrument), an expert based approach to item reduction may be preferable to a statistical approach.[2] This situation is far more likely in the patients' self assessment of symptoms. An expert based approach has been employed in very few studies that involved reducing indices, and mainly served to help choose among several solutions provided by statistical methods.[2] We chose the other route. We used an expert based reduction procedure, and statistical analyses of the quality of the items were secondary criteria. To reduce information bias in the reduction process, we combined two types of expert: patient experts, who had experience of the problems (representatives of the target population), and rheumatologist experts, using their knowledge of a broad spectrum of the disease.

The originality of our approach lies in the large number of experts involved in the study. Expert based approaches usually rely on the authors' own judgment of redundancy and insufficient face validity, or on the use of consensus methods with relatively small panels of experts. For instance, Guillemin and colleagues[12] used two panels when shortening the arthritis impact measurement scales 2 (AIMS2): one of 19 experts (rheumatologists, rehabilitation specialists, and methodologists) and another of 12 patients. Whitehouse *et al*[13] used a panel of 36 experts (orthopaedic surgeons, rheumatologists, nurses, physiotherapists, and research personnel). The large sample of patient experts ($n = 1218$) and rheumatologist experts ($n = 399$, approximately 15% of the rheumatologists in France) in our study is a good indicator of its representativeness and of the validity of the results.

The relevance of our reduction procedure is reinforced by the outcome of the procedure. The remaining items are the eight most important in the patients' opinion, and most of the excluded items were excluded in at least two steps of the reduction procedure. Taking patients' opinion into account in deriving short forms of validated questionnaires could improve the clinical relevance of such methods.

Whitehouse *et al*[13] proposed a seven item short form of the WOMAC function subscale, but the derivation process involved only a subgroup of patients with severe disease (patients undergoing hip or knee arthroplasty). In this context, the short form should be dedicated to assessing the outcome of total joint arthroplasty, as Whitehouse indicated. However, five items are shared between the Whitehouse form and our own. The particular population in Whitehouse's study may explain some of the discrepancies between the two short forms—especially that fact that activities such as "sitting" or "rising from bed," which are more likely to be impaired in severe disease, are two of the seven items included in Whitehouse's version but excluded from our version (because they were ranked 14th and 15th, respectively, by the patients).

The assessment of the performance characteristics of the WOMAC function subscale short form, its clinical relevance, and its acceptability require further studies in independent samples. Such studies should involve different versions of the WOMAC function subscale, as well as different language translations and different scaling formats, and should be conducted in different countries, in different clinical environments (for example, rheumatology, orthopaedic surgery, physiotherapy, rehabilitation), and with different interventions.

ACKNOWLEDGEMENTS

This study was supported by an unrestricted grant from Merck, Sharp & Dohme Chibret Laboratories, France.

Authors' affiliations

F Tubach, G Baron, P Ravaud, Département d'Epidémiologie, Biostatistique et Recherche Clinique, Groupe Hospitalier Bichat-Claude Bernard (Assistance Publique – Hôpitaux de Paris), Faculté Xavier Bichat (Université Paris 7), Paris, France
B Falissard, Faculté Paris XI, Département de santé publique, Hôpital Paul Brousse (Assistance Publique – Hôpitaux de Paris), Villejuif, France
I Logeart, Merck Sharp & Dohme, Chibret Laboratories, Paris, France
M Dougados, Service de Rhumatologie B, Hôpital Cochin (Assistance Publique – Hôpitaux de Paris), Faculté Paris V, Paris, France
N Bellamy, CONROD, Faculty of Health Sciences, The University of Queensland, Brisbane, Queensland, Australia

APPENDIX

Proposed WOMAC function subscale short form (eight items)

- Descending stairs
- Ascending stairs
- Rising from sitting
- Walking on flat
- Getting in/out of a car
- Going shopping
- Putting on socks/stockings
- Getting on/off the toilet

The WOMAC function subscale gradations in the Likert-scaled French Canadian 3.0 version are: 0 = none, 1 = slight, 2 = moderate, 3 = severe, 4 = extreme.

The WOMAC function subscale short form comprises a total of 32 possible points, with 0 being the best and 32 being the worst.

REFERENCES

1 **Moran LA**, Guyatt GH, Norman GR. Establishing the minimal number of items for a responsive, valid, health-related quality of life instrument. *J Clin Epidemiol* 2001;54:571–9.
2 **Coste J**, Guillemin F, Pouchot J, Fermanian J. Methodological approaches to shortening composite measurement scales. *J Clin Epidemiol* 1997;50:247–52.
3 **Streiner D**, Norman G. *Health measurement scales. A practical guide to their development and use*, 2nd ed. Oxford: Oxford University Press, 1995.
4 **Bellamy N**, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol* 1988;15:1833–40.
5 **McConnell S**, Kolopack P, Davis AM. The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC): a review of its utility and measurement properties. *Arthritis Rheum* 2001;45:453–61.
6 **Altman R**, Alarcon G, Appelrouth D, Bloch D, Borenstein D, Brandt K, *et al*. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. *Arthritis Rheum* 1991;34:505–14.
7 **Altman R**, Asch E, Bloch D, Bole G, Borenstein D, Brandt K, *et al*. Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Association. *Arthritis Rheum* 1986;29:1039–49.
8 **Choquette D**, Bellamy N, Raynauld J. A French-Canadian version of the WOMAC osteoarthritis index. *Arthritis Rheum* 1994;37(S226).
9 **Cronbach LJ**. Coefficient alpha and the internal structure. *Psychometrika* 1951;16:297–334.
10 **Coste J**, Fermanian J, Venot A. Methodological and statistical problems in the construction of composite measurement scales: a survey of six medical and epidemiological journals. *Stat Med* 1995;14:331–45.
11 **Felson DT**, Zhang Y. An update on the epidemiology of knee and hip osteoarthritis with a view to prevention. *Arthritis Rheum* 1998;41:1343–55.
12 **Guillemin F**, Coste J, Pouchot J, Ghezzail M, Bregeon C, Sany J. The AIMS2-SF: a short form of the Arthritis Impact Measurement Scales 2. French Quality of Life in Rheumatology Group. *Arthritis Rheum* 1997;40:1267–74.
13 **Whitehouse SL**, Lingard EA, Katz JN, Learmonth ID. Development and testing of a reduced WOMAC function scale. *J Bone Joint Surg Br* 2003;85:706–11.

SECTION 8 - Responder Criteria and State-Attainment Criteria

The traditional approach to the analysis of data from OA clinical trials has been performed at the group level. More recently, attention has focused on individual patient reported outcomes. These can be considered as being of two general forms: responder criteria in which each patient is classified as a responder or non-responder to treatment, based on whether their change in health status exceeds a pre-defined threshold, and state-attainment criteria in which patients are classified, not on the basis of change (better, same, worse), but on the basis of when, whether, and/or for how long they achieve a certain pre-defined level of low symptom severity. Research in both areas is developmental and can be considered proposals, rather than guidelines or requirements. Nevertheless, responder criteria and state-attainment criteria represent innovations in the analysis of clinical trial outcomes, and may provide a bridge between health outcomes assessment strategies in clinical research and clinical practice environments.

Although responder criteria had been proposed by the American College of Rheumatology in 1995 for clinical trials in patients with rheumatoid arthritis, no comparable development had occurred for clinical trials in patients with OA. Following the development of consensus on core set clinical measures for OA clinical trails (22,23), a Task Force of the Osteoarthritis Research Society International engaged in the development of response criteria for clinical trials in hip and knee OA (35). This initiative, in part based on WOMAC data, resulted in the specification of two sets of criteria (Propositions A and B), which were joint and intervention-class specific, and of defined levels of sensitivity and specificity (35). The OARSI Response Criteria are based on minimum thresholds, for combined absolute and relative change, on one or more of the core set clinical variables (pain, function, patient global assessment).

The relationship between changes registered on the WOMAC Index and the same patients' perceived global assessment of their response to therapy was explored in the context of data from two identical RCTs involving rofecoxib (36). The analyses determined that the Minimum Perceptible Clinical Improvement (MPCI) for the WOMAC VA3.0 Index subscales in patients with hip and knee OA, was 9.7 mm (0-100 scale) for pain, 9.3 mm for function and 10 mm for stiffness (36). The study suggested that the MPCI might provide a better assessment of the clinical relevance of the effects of therapeutic interventions in OA (36).

The OMERACT V Conference provided an opportunity to further explore issues relating to the measurement of change, definitions of patient response and responder criteria in OA (37,38). Various definitions of change were discussed including the Minimum Change Potentially Detectable (MCPD), the Minimum Percentage Change Potentially Detectable (MPCPD), the MPCI, the Minimum Clinically Important Difference (MCID), the minimal change detectable beyond measurement error, OARSI responder criteria, and patient perceptions of change (37). Consensus-based estimates, acquired through interactive touch-pad polling of delegate opinion, indicated support for the following: a) Development of response criteria for individuals with other diseases, b) The importance of defining major clinically important improvement for RA, c) The importance of validating short-term response/improvement in predicting long- term outcome, and d) Defining response criteria for OA in terms of both percent and absolute

change (38). In essence the OMERACT consensus was supportive of the general approach taken by the OARSI Task Force in developing responder criteria for OA clinical trials (35,38).

In an attempt to further explore the original OARSI responder criteria (Propositions A and B), and potentially simplify the presentation of the OARSI responder criteria, a collaborative exercise was undertaken between OMERACT and OARSI (39). Six scenarios (Original Propositions A and B, and four new Propositions C-F) were evaluated, in what were termed "elaboration" and "revisit" data sets (39). The majority of the pain evaluations and almost all the functional evaluations, in contributing studies, had been performed with the WOMAC Index (See Table 1 in Reference 39). The initiative resulted in the proposal of a single international simplified set of responder criteria for OA clinical trials, that were applicable to hip and knee OA, and independent of intervention class (39).

In order to further explore the application of responder criteria, two secondary analyses (40,41) of the previously published pharmaco-economic evaluation of hylan G-F 20 (17,18) were undertaken. The first analysis paralleled developments in RA, where ACR 20, 50 and 70 levels of response had already been investigated in RA patients. The evaluation of the WOMAC 20, 50, 70 response criteria in knee OA patients, based on the WOMAC LK3.0 Index, provided preliminary evidence, supporting the capacity of WOMAC 20,50 and 70 responder criteria, to detect clinically important, statistically significant between-group differences in a pragmatic randomised trial (40). The second analysis was performed to assess the performance of the OARSI Responder Criteria (Propositions A and B), and the OMERACT-OARSI Response Criteria (Proposition D) in knee OA patients (41). The analyses provided support for the capacity of all three criteria to detect clinically important statistically significant differences between treatment groups, in an RCT of a hylan class intervention (41).

The aforementioned criteria are largely based on a combination of expert opinion and statistical techniques. In order to incorporate the patient's own perspective, a study was conducted, through collaboration with colleagues in France, and resulted in preliminary definitions for the Minimally Clinically Important Improvement (MCII)(42), and the Patient Acceptable Symptom State (PASS) (43) in knee and hip OA. The definitions for physical function were entirely based on data captured by the WOMAC 3.0 Index. The following definitions for MCII absolute on 0-100 scales (and relative %) changes for knee and hip OA respectively, were suggested by the analyses: a) Pain –19.9 mm (-40.8%) and –15.3 mm (-32.0%); b) patient global assessment –18.3 mm (-39.0%) and –15.2 mm (-32.6%); and c) WOMAC physical function –9.1 mm (-26.0%) and –7.9 mm (-21.1%) (42).

The Patient Acceptable Symptom State (PASS) is a novel concept, and is the 75th percentile of the symptom severity score of patients who consider their health state to be satisfactory. The following definitions for PASS (threshold values on 0-100 scales) for knee and hip OA, respectively, were suggested by the analyses: a) Pain 32.3 mm and 35.0 mm; b) patient global assessment 32.0 mm and 34.6 mm; and c) WOMAC physical function 31.0mm and 34.4 mm (43). The development of state-attainment criteria, in OA, is very new, and the analyses provide preliminary information (43). It is acknowledged that the concept requires further investigation and the propositions requiring further validation (43). Nevertheless, patient involvement in estimating the clinical importance of

improvement (42) and the acceptability of different levels of symptom severity (43) is innovative, meets the obligations and requirements for consumer involvement in decision making and establishes preliminary consumer-based definitions for response and state attainment in knee and hip OA.

References:

35. Dougados, M., LeClaire, P., van der Heijde, D., Bloch, D.A., Bellamy, N., and Altman, R.D. Special Article: Response criteria for clinical trials on osteoarthritis of the knee and hip: A report of the Osteoarthritis Research Society International Standing Committee for Clinical Trials Response Criteria Initiative. Osteoarthritis and Cartilage 2000;8(6):395-403

36. Ehrich E.W., Davies G.M., Watson D.J., Bolognese J.A., Seidenberg B.C., and Bellamy N. Minimal Perceptible Clinical Improvement with the Western Ontario and McMaster Universities osteoarthritis index questionnaire and global assessments in patients with osteoarthritis. Journal of Rheumatology 2000;27(11):2635-2641.

37. Bellamy, N., Carr, A., Dougados, M., Shea, B., and Wells, G. Towards a definition of "difference" in osteoarthritis. Journal of Rheumatology 2001;28(2):427-430.

38. Wells G., Anderson J., Beaton D., Bellamy N., Boers M., Bombardier C., Breedveld F., Carr A., Cranney A., Dougados M., Felson D., Kirwan J., Schiff M., Shea B., Simon L., Smolen J., Strand V., Tugwell P., van Riel P., and Welch V.A. Minimal Clinically Important Difference Module: Summary, Recommendations, and Research Agenda. Journal of Rheumatology 2001;28(2):452-454.

39. Pham T., van der Heijde D., Altman R.D., Anderson J.J., Bellamy N., Hochberg M., Simon L., Strand V., Woodworth T. and Dougados M. OMERACT-OARSI Initiative: Osteoarthritis Research Society International set of responder criteria for osteoarthritis clinical trials revisited. Osteoarthritis and Cartilage 2004;12:389-399.

40. Bellamy N., Bell M., Goldsmith C., Pericak D., Walker V., Raynauld J-P., Torrance G., Tugwell P. and Polisson R. Evaluation of WOMAC 20, 50, 70 Response Criteria in patients treated with hylan G-F 20 for knee osteoarthritis. Annals of the Rheumatic Diseases. 2005;64:881-885.

41. Bellamy N., Bell M.J., Goldsmith C.H., Pericak D., Walker V., Raynauld P-J., Torrance G.W., Tugwell P. and Polisson R. The effectiveness of Hylan G-F 20 in patients with knee osteoarthritis: an application of two sets of response criteria developed by the OARSI and one set developed by OMERACT-OARSI. Osteoarthritis and Cartilage 2005;13:104-110.

42. Tubach F., Ravaud P., Baron G., Falissard B., Logeart I., Bellamy N., Bombardier C., Felson D.T., Hochberg M., van der Heijde D. and Dougados M. Evaluation of clinically relevant states in patient-reported outcomes in knee and hip osteoarthritis: The Minimal Clinically Important Improvement. Annals of the Rheumatic Diseases. 2005;64(1):29-33.

43. Tubach F., Ravaud P., Baron G., Falissard B., Logeart I., Bellamy N., Bombardier C., Felson D.T., Hochberg M., van der Heijde D. and Dougados M. Evaluation of clinically relevant states in patient-reported outcomes in knee and hip osteoarthritis: The Patient Acceptable Symptom State. Annals of the Rheumatic Diseases. 2005;64(1):34-37.

OsteoArthritis and Cartilage (2000) 8, 395–403
© 2000 OsteoArthritis Research Society International
doi:10.1053/joca.2000.0361, available online at http://www.idealibrary.com on IDEAL®

1063–4584/00/060395+09 $35.00/0

Osteoarthritis and Cartilage
Journal of the OsteoArthritis Research Society International

OARSI OSTEOARTHRITIS RESEARCH SOCIETY INTERNATIONAL

Special article: Response criteria for clinical trials on osteoarthritis of the knee and hip

A report of the Osteoarthritis Research Society International Standing Committee for Clinical Trials Response Criteria Initiative

M. Dougados*, P. LeClaire*, D. van der Heijde†, D. A. Bloch‡, N. Bellamy§ and R. D. Altman||
*Institut de Rhumatologie, Hôpital Cochin, 27 rue du Faubourg Saint Jacques, 75014 Paris, France
†Department of Rheumatic Diseases, University Hospital Maastricht, PO Box 5800, 6202 AZ Maastricht, The Netherlands and Limburg University Center, Diepenbeek, Belgium
‡Department of Health Research and Policy, HRP Redwood Building, Stanford University, Stanford, California 94305-5405, U.S.A.
§Centre of National Research on Disability and Rehabilitation Medicine, The University of Queensland, Royal Brisbane Hospital, Herston Road, 4029 Queensland, Australia
||Miami Veterans Affairs Medical Center and University of Miami, Miami, Florida 33125, U.S.A.

Summary

Background: The domains of pain, function and patient's global assessment are identified as core variables and frequently measured in clinical trials of patients with osteoarthritis (OA) of the hip and knee.

Objective: To develop response criteria for OA of hip and knee based on the domains of pain, function and patient's global assessment.

Methods: A methodology was developed by an interaction of the Osteoarthritis Research Society International Standing Committee on Clinical Trials, biostatisticians, pharmaceutical company representatives and health agency representatives. Data from previously conducted placebo-controlled clinical trials were normalized and collated. Data were subset by location of OA (knee, hip), active agent used in the clinical trial (non-steroidal anti-inflammatory drug, other agent) and route of administration (oral, intra-articular). Statistical analysis identified response criteria which best discriminate active agent from placebo.

Results: Based on the analysis of data from 14 studies (totaling 1686 patients) and consensus opinion, the optimal responder criteria set differed for location of OA, active agent to be used, and route of administration. Because of nearly identical statistical results, two sets of responder criteria are proposed: (1) 'high' pain response or, alternatively, a 'moderate' response for at least two of three domains: pain, function and patient's global assessment; (2) 'high' response for either pain or function or, alternatively, a 'moderate' response for at least two of three domains: pain, function and patient's global assessment. The sensitivity (i.e., the percentage of responders in the active group) ranged from 52 to 96% and the specificity (i.e., the percentage of nonresponders in the control group) from 47 to 73%.

Conclusion: Based on data from clinical trials, two sets of responder criteria have been developed that can categorize an individual's responses to treatment in a clinical trial. These responder criteria require validation in additional datasets. © 2000 OsteoArthritis Research Society International

Key words: Clinical Trials Response Criteria Initiative.

Introduction

Presently available therapeutic modalities for osteoarthritis (OA) are directed at symptoms, with no device or drug consistently shown to modify structure (joint pathology). Definitions for the design and conduct of clinical trials have been developed by the Osteoarthritis Research Society International (OARSI)[1] the World Health Organization (WHO), the International League of Association for Rheumatologists (ILAR),[2] the experts involved in Outcome Measures in Arthritis Clinical Trials (OMERACT),[3] and the European Group for the Respect of Ethics and Excellence in Sciences (GREES) through an Osteoarthritis Subcommittee.[4] All recommend a clear separation of the evaluation of the symptoms from the evaluation of the structure (disease) in OA.

In the evaluation of symptoms of OA, several domains can be considered. Examples include pain, function, inflammation, range of motion, quality of life, patient's global opinion, physician's global opinion, etc. However, discussion at OMERACT III focused on three domains, identified as core variables to be included in all clinical

Address correspondence to: Roy D. Altman, M.D., NH207G, Miami VAMC, 1201 NW 16th Street, Miami, FL 33125, U.S.A. Tel.: +1 305-243-5735; Fax: +1 305-243-5655; E-mail: raltman@med.miami.edu

studies involving OA, i.e. pain, function and patient's global assessment.

Within each of these three domains, several instruments may be considered, such as a simple visual analog scale (VAS) or Likert scale. Measurement may also be by a more complex instrument, such as the Lequesne Functional Severity Index,[5] the Western Ontario McMasters Universities Osteoarthritis (WOMAC) index,[6] the Health Assessment Questionnaire (HAQ) and the Arthritis Impact Measurement Scales (AIMS).[7]

While defining a core set of domains to be measured represents an advance in defining and standardizing the conduct of OA clinical trials, the analysis and reporting of results continues to be based on average improvement for the study population on each of the outcomes measured. Hence, the average improvement experienced by a group of treated patients is compared to the average improvement of another group of treated patients (treatment of which may be placebo). In contrast, for rheumatoid arthritis there have been at least two sets of multi-variable response criteria proposed in order to present the results obtained in clinical trials for an individual.[8,9] There are presently no such criteria available for OA.

Such response criteria in clinical trials offer the following advantages:

(1) They permit a single statistical analysis. This takes into account clinically important changes on multiple variables, without the need for correction for multiple comparisons, and provides an exact determination for each individual as to whether or not they have responded to the study treatment.
(2) They facilitate the categorical analysis of groups of patients in clinical trials based on definitions of responders and non-responders.
(3) They allow the comparison of data from different clinical trials in OA.

Response criteria may also have application in the calculation of sample size for clinical trials,[10] assist in pharmacoeconomic analysis and in data analysis of the Cochrane Collaborative Project,[11] and facilitate calculation in the number needed to treat approach of data analysis.[12] For these reasons, the OARSI Standing Committee on Clinical Trials elected to develop a set of responder criteria for the knee and hip. The committee felt that any such criteria should be research-based to the extent that is possible.

Methods

BASIC APPROACH

There are several potential approaches to the development of response criteria. One utilizes data from clinical trials with a statistical strategy that defines the criteria. This approach is limited by the quality of the data, but has the advantage of being evidence-based. Another approach relies solely on the judgement of experts. The Delphi methodology may help focus the process. This approach is limited because the developed criteria may be statistically unobtainable, but has the advantage that they probably represent clinically relevant goals of management. The OARSI Standing Committee on Clinical Trials elected to take a third approach, one that takes advantage of the strengths of both strategies, while minimizing the impact of their disadvantages.

The Standing Committee on Clinical Trials first set out to clearly define the process for the development of response criteria. Although the majority of the effort would be by the committee members, different steps of the project would be discussed among experts from several fields: rheumatologists, epidemiologists, biostatisticians, representatives of the pharmaceutical industry and representatives of health regulatory agencies [Food and Drug Administration (FDA) and European Agency for the Evaluation of Medicinal Products (EMEA)].

It was agreed that the data to be processed in the analysis would be provided by the pharmaceutical industry, and that these data would be from clinical trials performed under their direction by good clinical practice guidelines. The clinical trial data would be from placebo controlled clinical trials involving an active agent of a minimum 6 week duration of study. The active agent would not be identified by name, but only by class of agent. In order to maintain the anonymity of the agents, the contributors to this trial are not identified in the text or the acknowledgement. Although not specified in the data collection, all agents have been approved by either the European or US health regulatory agencies.

DATASET COLLECTION

Datasets and protocols from 37 clinical trials were provided by the participating pharmaceutical and biotechnology companies. To be eligible, the datasets were to include double-blind, parallel and placebo controlled trials conducted on patients suffering from either hip or knee OA in which at least one of the following clinical variables was collected as a primary efficacy variable during the trial: pain or function (as an assessment of functional disability), or a global assessment by the patient. Minimal data had to include these variables at baseline and after treatment. Studies included in the analysis had to show a benefit (positive study) from the study medication/agent when compared to placebo for the primary efficacy variable ($P \geq 0.05$). The study did not include non-pharmacologic modalities.

DEFINITION OF THE OUTCOMES WITHIN EACH DOMAIN

Studies often applied different outcome assessment instruments in the measurement of each of the domains. For example, pain may have been measured by the five-question subscale of the WOMAC, as a part of the Lequesne Functional Severity Index, or by a single question. The following outcomes were chosen for subsequent analyses.

Pain

The pain measurement selected for analysis included the following in rank order: (1) global measure of pain; if not available, (2) WOMAC pain subscale; if not available, (3) average of the four questions of Lequesne Functional Severity Index that focused on pain (i.e. pain at night, pain in the standing position, pain during walking and pain while switching from a sitting to a standing position).

Function

The functional disability measurement selected for analysis included the following in rank order: (1) WOMAC

functional subscale; if not available, (2) average of the four questions of Lequesne Functional Severity Index that focused on functional impairment (i.e. the last four questions of either the hip or knee Lequesne Functional Severity Index).

Some datasets contained only the total score of the Lequesne Functional Severity Index, while others contained both the total score and the response to each of the 11 questions. We examined the correlation of the information contained in the 11 questions of the Lequesne functional severity index, and the four specific questions on function. This analysis was conducted on available baseline data from 1748 patients. The following equation was able to explain >63% of the variance:

$$\text{Lequesne's function subscale} = -0.22569 + 0.43204 \times (\text{Lequesne total}) - 0.00691 \times (\text{Lequesne total})^2$$

This methodology was implemented in one study of 342 patients.

Patient's global assessment

The patient's global assessment measurement was selected for analysis. Because not all data sets contained this information, we attempted to explain the variable evaluating 'patient's global assessment' by the information contained in the variables evaluating the domains 'pain' and 'function'. Analysis of 892 patients' data, from five studies where all three domains were available, revealed that pain and function explained no more than 27% of the variance of the patient's global assessment. Hence, studies without a patient's global assessment could not be used in the development of the responder index.

RESCALING OF THE OUTCOME MEASURES

Studies often utilized instruments with different scales. For example, some studies have used a 10 cm VAS or a Likert scale with a variable number of questions. Other examples include a 0–8 interval for the pain questions from the Lequesne functional severity index and the 0–500 interval for the sum of the pain subscale of the WOMAC. When both were present in the same study, a 100 mm VAS was used in lieu of a Likert scale. All scales that used anything other than a 0–100 measurement were rescaled to a 0–100 interval for the purposes of this study, henceforth referred to as normalized units or NU.

FORMAT OF THE SET OF CRITERIA

The following three formats were considered:

1. The 'Osteoarthritis Disease Activity Score'.[13] This format involves a composite index using an equation obtained using multiple regression. The advantage of this method is that it develops a single number summarizing the clinical symptoms for each single patient at each visit. However, its use is usually based on a discriminate analysis comparing high and low disease activity. Such analysis requires an external judgement, which was not available in the different datasets. Therefore, this proposition was rejected.

2. The 'tree' type format.[14] In this format, patients are partitioned into mutually exclusive categories based on the outcome measures. The tree selects the most appropriate variable that can differentiate two populations. It does the selection in sequential steps, selecting the most appropriate variable at each step. The final tree allows classification of all patients as responder or nonresponder.

3. The American College of Rheumatology rheumatoid arthritis, or 'ACR-RA-set' format.[8] This technique counts the number of criteria present and if that number exceeds some pre-specified number the patient is classified as a responder. This format is easily understood by clinicians and easily applied in practice. This method does not consider the severity of disease activity at baseline. For example, a 50% improvement can be obtained in a patient with a pain VAS decreasing from 80 to 40, but also for a patient with a pain VAS decreasing from 6 to 3.

It was elected to take advantage of the latter two formats with the addition of a minimum response. For example, an 'X'% improvement in pain, function or patient's global assessment from baseline would have to be accompanied by a minimum improvement of 'Y' NU (rescaled variable in normalized units, see above). Intent-to-treat analysis (ITT) with last observation carried forward (LOCF) was used for all studies.

STUDIED POPULATIONS

Data were available to allow the separation of the following subsets: (1) patients with knee OA on an oral non-steroidal anti-inflammatory drug (NSAID); (2) patients with knee OA on a non-NSAID anti-OA oral drug; (3) hip OA on an oral NSAID; (4) knee OA receiving an intra-articular anti-OA specific non-steroidal agent. Insufficient data were available to examine hip OA on a non-NSAID anti-OA drug, intra-articular depocorticosterolds and intra-articular agents for hip OA.

PROPOSED SET OF CRITERIA

Eighteen scenarios defining responders were examined. To illustrate a very conservative scenario, a patient is considered as a responder if he/she fulfills the following: 'an improvement in pain AND in function AND in global assessment'. For each proposed scenario, a statistical analysis based on the response criteria capacity to discriminate active and placebo groups was used to evaluate the performance of different cut-off points (in both the percentage of change and the absolute change) for each single variable included in this scenario. However, before this analysis, the members of the Standing Committee on Clinical Trials proposed a range for each variable that should be explored together with a minimum interval between the different values that should be evaluated (i.e. a minimum interval of 5% or 5 NU). Different definitions of response criteria corresponding to different combinations of cut-offs were evaluated. The definition retained (with fixed cut-offs) was the one which maximizes the difference between the percentage of responders in the active group ('sensitivity') and the percentage of responders in the placebo group ('1-specificity').

Results

PATIENTS AND STUDY COURSE

The number of studies examined and the screening of those studies are illustrated in Fig. 1. The majority of the

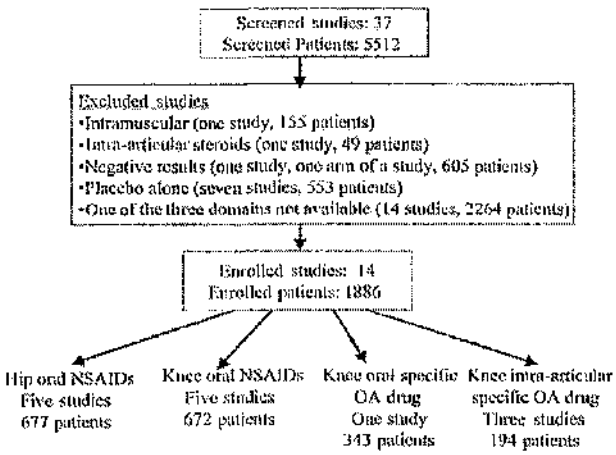


Fig. 1. Flow diagram for numbers of studies and patients.

excluded studies were because of absence of at least one of the domains to be examined, mostly the patient's global assessment. Figure 1 also outlines the number of patients involved in the screening process and the numbers derived for each of the subsets to be examined. Although requested, there were no trials available to examine analgesics (narcotic or non-narcotic) in OA. The screening reduced the number of trials from 37 to 14 and the number of patients in those trials from 5512 to 1886. The majority of information was on NSAIDs for knee and hip.

The characteristics of the study populations are summarized in Table I. The placebo and active drug groups were similar in age, sex and body mass index. The populations were relatively overweight as reflected by a high body mass index. In general there were more patients receiving active drug than placebo as several trials had more than one arm receiving active drug. The trials examined were relatively recent as reflected by the common use of the WOMAC.

The baseline values in NU and the percent change during the study of the three selected domains are summarized for each of the study subsets in Table II. These data confirm that most of the patients were quite symptomatic at entry since the mean values of pain, functional impairment and patient's global assessment were all over 40 NU. Moreover, the values obtained in the variable 'patient's global assessment' were over 60 NU, higher than the values for 'pain' and/or 'function'. Consistent with the requirement that only positive studies be included, the changes from baseline in the 'active' group were of greater magnitude than the 'placebo' group for all domains. It was not possible to determine worsening in the studies of knee intra-articular specific anti-OA drug, as the the change in 'function' and 'patient's global assessment' were collected at the end of the study by asking the patients about the level of his/her 'improvement'.

DEVELOPMENT OF SETS OF CRITERIA

There were 18 different response criteria evaluated. Each was evaluated for NSAID hip, NSAID knee, anti-OA drug hip, anti-OA drug knee and knee intra-articular specific anti-OA drug. Numerous cut-offs on percent change and change in NU were reviewed on each of the 18 response criteria for each of the three domains. Each evaluation included sensitivity and specificity. Two sets of response criteria were statistically superior and felt to be clinically relevant (Figs 2 and 3).

The first scenario (proposition A) (Fig. 2) emphasizes the domain 'pain'. A 'high' improvement in pain was sufficient to define a responder. However, using this set of criteria, a patient can be also considered as a responder if an improvement of 'moderate' magnitude is observed in two of

Table I
Characteristics of the patients included for the development of the response criteria

Characteristics	Selected population	
	Placebo	Active drug
Age (years) (mean±S.D.)	63±10	63±10
Sex (% women)	72	72
Body mass index (kg/m^2) (mean±S.D.)	31± 7	30± 7
Localization		
Hip (number)	197	480
Knee (number)	362	847
Route of administration		
Oral (number)	460	1232
Intra-articular (number)	99	95
Class of the active drug		
NSAIDs (number)	372	977
Specific anti-OA drug (number)	187	350
Domains and outcome assessment instrument		
Pain		
Global VAS pain (number, 3 studies)	179	563
WOMAC pain subscale (number, 8 studies)	281	658
4 questions of the algo-functional index (number, 3 studies)	99	95
Function		
WOMAC pain subscale: (number, 8 studies)	281	658
4 questions of the algo-functional index (number, 3 studies)	179	574
Patient's global assessment		
Global VAS (number)	228	707
Global Likert scale (number)	232	525

Table II

Baseline values and changes after therapy (percentage±standard deviation: standardized units, 0–100 scale) in the three selected domains (pain, function, patient's global assessment) with regard to osteoarthritis localization and category of the active drug

Domains	Hip osteoarthritis NSAID study therapy		Knee osteoarthritis NSAID study therapy		Knee osteoarthritis specific therapy		Knee osteoarthritis intra-articular specific therapy	
	Placebo *N*=197	Active drug *N*=480	Placebo *N*=175	Active drug *N*=197	Placebo *N*=88	Active drug *N*=255	Placebo *N*=99	Active drug *N*=95
Pain								
Baseline value	64±19	62±21	58±22	59±22	58±12	59±12	56±20	52±19
Changes during the study	−14±53	−34±62	−4±73	−23±90	−34±55	−46±43	−24±40	−61±28
Functional impairment								
Baseline value	58±17	58±16	56±20	56±21	41± 7	41± 8	NA	NA
Changes during the study	−2±38	−21± 4	−7±64	−23±54	−28±45	−34±41	−42±16	−47±10
Patient's global assessment								
Baseline value	72±16	72±17	70±18	72±17	61±15	61±16	NA	NA
Changes during the study	−22±59	−45±40	−19±40	−41±39	−30± 5	−41±43	−39±32	−79±26

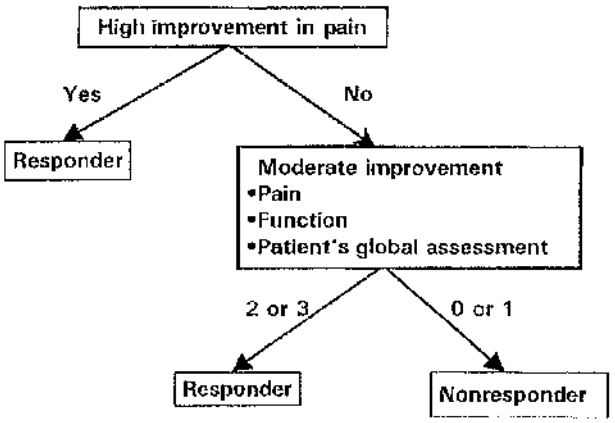


Fig. 2. OARSI Responder Criteria—Proposition A. Decision tree is utilized for each patient. If there is a 'high' improvement in pain, the person is considered a responder. If there is not a 'high' response to pain, they must have a 'moderate' response to two or three of the domains to be labeled a responder.

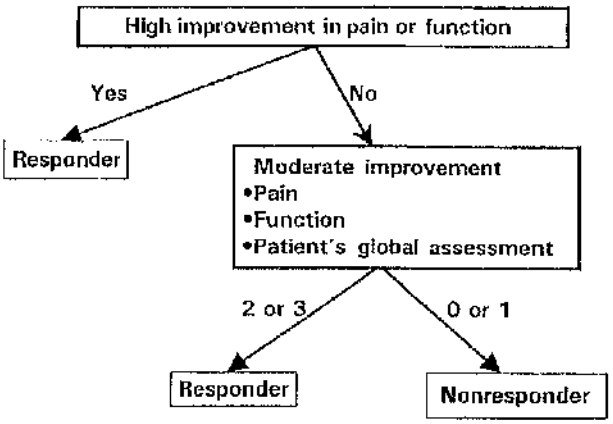


Fig. 3. OARSI Responder Criteria—Proposition B. Decision tree is utilized for each patient. If there is a 'high' improvement in pain or function, the person is considered a responder. If there is not a 'high' response to pain, they must have a 'moderate' response to two or three of the domains to be labeled a responder.

the three domains, i.e. pain, function and patient's global assessment.

The second scenario (proposition B) (Fig. 3) is similar to the first and with nearly identical statistical results. This scenario applies equal importance to 'pain' and 'function', requiring a 'high' response of one OR the other. Alternatively, a 'moderate' magnitude of response could be present in two of the three domains.

Optimal cut-offs for the different subgroups were proposed based on statistical analysis (Tables III and IV). The proposed cut-offs are different for proposition A and proposition B and are different for each intervention and joint.

Attempts to develop a single OA response criteria that would include the different subgroups led to important loss of separation of active and placebo groups; the loss was only 1–2% in sensitivity, but nearly 15% in specificity.

Considering the first step of the decision tree for both sets of response criteria, the different statistical analyses in the different subgroups concluded that an improvement of at least 40% was required (ranging from 40 to 60%) together with an absolute improvement of at least 20 NU ranging from 20 to 30. Considering the second step, an improvement of lower magnitude was required—a relative improvement ranging from 15 to 35% and an absolute improvement ranging from 10 to 20 NU. The results concerning the performance of the sets of response criteria in terms of sensitivity (i.e. percentage of responders in the group of patients who received the active drug) and 1-specificity (i.e. percentage of responders in the group of patients who received the placebo) are summarized in Tables V and Vi.

Figure 4 exemplifies how the tables can be used and corresponds to the second row of Table III: responder criteria for knee OA using an oral NSAID. A major criterion would be a 45% reduction in pain with a minimum decrease of 20 NU (i.e. 20 mm on a 100 mm VAS). If that were not achieved, the minor criterion for a responder would be someone achieving a change in at least two of the following three domains: a reduction of pain of at least 15% (minimum reduction of 10 NU); improvement in function of at least 30% (minimum improvement of 15 NU); and/or an improvement in patient's global assessment of at least 35% (minimum improvement of at least 10 NU).

In the NSAIDs trials (hip or knee), the percentage of responders is close to what might be expected from prior trials in rheumatic diseases (i.e. 50–60% in the active

Table III
Optimal cut-offs to be applied for the OARSI-Responder Criteria—proposition A (see Fig. 2)

Subgroup	High improvement in pain		Moderate improvement in Pain		Function		Global assessment	
	Relative change*	Absolute change**	Relative change	Absolute change	Relative change	Absolute change	Relative change	Absolute change
Hip NSAIDs	60	20	35	20	20	10	30	10
Knee oral NSAIDs	45	20	15	10	30	15	35	10
Knee oral specific drug	55	30	35	10	15	20	15	15
The three above groups together	55	30	35	15	15	20	15	15
Knee intra-articular specific drug	40	30	35	15	35	10	30	10

*Relative change: percentage of change during the study (final minus baseline over baseline×100).
**Absolute change: absolute change during the study (final minus baseline on a 0–100 interval scale).

Table IV
Optimal cut-offs to be applied for the OARSI-Responder Criteria—proposition B (see Fig. 3)

Subgroup	High improvement in Pain		Function		Moderate improvement Pain		Function		Global assessment	
	Relative change*	Absolute change**	Relative change	Absolute change	Relative change	Absolute change	Relative change	Absolute change	Relative change	Absolute change
Hip NSAIDs	50	30	50	20	25	15	20	10	20	10
Knee oral NSAIDs	50	20	60	20	30	15	20	20	25	10
Knee oral specific drug	55	30	50	20	30	20	20	20	20	15
The three above groups together	55	30	50	20	30	15	20	20	20	15
Knee intra-articular specific drug	50	30	60	20	20	20	30	10	30	10

*Relative change: percentage of change during the study (final minus baseline over baseline×100).
**Absolute change: absolute change during the study (final minus baseline on a 0–100 interval scale).

Table V
Percent of study patients meeting OARSI responder criteria—proposition A

Subgroup	High improvement in pain		Moderate improvement in 2 of the 3: pain, function, global assessment		Total	
	Percentage responders		Percentage responders		Percentage responders	
	Active*	Placebo**	Active	Placebo	Active	Placebo
Hip NSAIDs	35%	18%	27%	14%	62%	33%
Knee oral NSAIDs	39%	19%	13%	8%	52%	27%
Knee oral specific drug	49%	39%	13%	12%	62%	51%
Knee intra-articular specific drug	44%	18%	48%	29%	92%	47%

*Active: sensitivity=% responders on active drug.
**Placebo: 1—specificity=% responders on placebo.

groups vs 20–30% in the placebo group). These figures are different in both the 'knee systemic specific drug' and the 'knee intra-articular specific drug' in which the percentage of placebo responders was higher (respectively, 51 and 47% in proposition A and 50 and 47% in proposition B).

Discussion

This study combined the efforts of academic researchers, biostatisticians, representatives of the pharmaceutical industry and representatives of health agencies to develop responder criteria for clinical trials of OA of hip and knee. The data led the Standing Committee on Clinical Trials to propose two sets of responder criteria, with variation in the specific recommendations for hip and knee and for different therapeutic approaches.

The authors are unaware of any previously proposed responder criteria for OA. The proposed responder criteria include the core set of outcome measures for OA clinical trials developed at OMERACT III: pain, function and patient's global assessment. These outcome measures have also been recommended by the OARSI and various health regulatory agencies.

There are limitations to this study. Because of variations in study design a majority of the screened studies, and

Table VI
Percent of study patients meeting OARSI responder criteria—proposition B

Subgroup	High improvement in pain		Moderate improvement in 2 of the 3: pain, function, global assessment		Total	
	Percentage responders		Percentage responders		Percentage responders	
	Active*	Placebo**	Active	Placebo	Active	Placebo
Hip NSAIDs	39%	27%	24%	12%	69%	39%
Knee oral NSAIDs	39%	19%	12%	7%	51%	26%
Knee oral specific drug	52%	40%	9%	10%	61%	50%
Knee intra-articular specific drug	51%	29%	39%	18%	91%	47%

*Active: sensitivity=% responders on active drug.
**Placebo: 1—specificity=% responders on placebo.

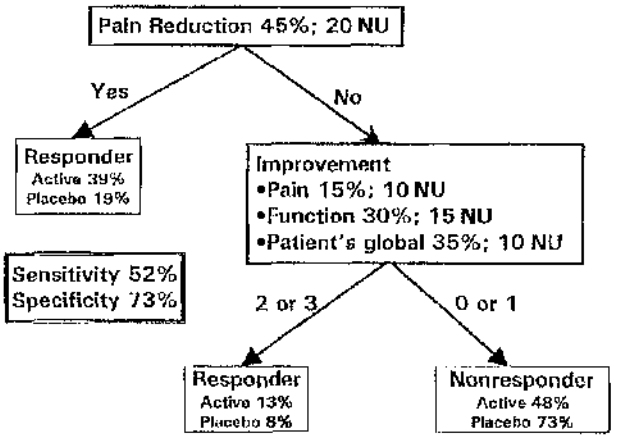


Fig. 4. OARSI Responder Criteria—example of proposition A for knee oral NSAID subgroup. This decision tree exemplifies the use of the algorithm for a patient in a clinical trial that is testing an non-steroidal anti-inflammatory drug for osteoarthritis of the knee. In this instance a 'high' improvement in pain equates to a 45% reduction in pain that is also a minimal change of 20 NU or normalized units. Twenty normalized units equates to a minimal change of 20 mm on a 100 mm scale. If the patient achieves this reduction in pain, they are considered a responder. If they do not, to be considered a responder the patient must achieve a positive result in at least two of the three domains: i.e. pain would need a reduction of 15% with a minimal change of 10 mm on a 100 mm scale, function would require a 30% improvement for a minimal 15 mm on a 100 mm scale and the patient's global assessment would need a 35% improvement with a minimal 10 mm change on a 100 mm scale. This algorithm labeled 52% of the 197 patients on an NSAID as responders and 73% of the 175 patients on placebo as non-responders.

hence a majority of the patients, had to be excluded from the analysis. Even in studies included in the analysis scores had to be generated from existing data that had been collected in different formats. The above studies were of variable duration; the influence of time of study on the response criteria was not addressed.

Different information gathering instruments were used. The outcomes from these different instruments may be similar but their responsiveness to change may be different. There is a question as to whether subscales can be extracted from instruments that were not developed around a subscale structure. There are also limitations imposed by modeling in order to input values of real data. Few, if any of the studies asked for the patient's global assessment in the same way.

The different instruments and even some of the same instruments used different scales, requiring a rescaling to a common 0–100 scale in order to normalize values. It is recommended that future studies use a 0–100 interval. This would facilitate analysis and provide uniform and understandable communication. The 0–100 interval would allow the use of the proposed responder criteria, emphasizing minimum change in NU.

The proposed responder criteria utilize a format that requires an absolute change. The technique also includes a second layer in a simple 'tree' format giving heirarchal application of criteria. In order to address severity of disease at baseline, a minimum level of improvement was also required. Whatever the rating scale used, the application of the proposed set of response criteria is feasible (for example, a required absolute improvement of 10% in the variable means an improvement of at least one grade for a 0–4 Likert scale, 2.4 points for a 0–24 scale, etc.).

It should be pointed out that the use of these responder criteria will require a protocol to set minimum entry criteria, since attainment of response by these criteria must remain an achievable target for all participants. Individuals with baseline values less than the minimum required absolute change can never be designated as responders, even if they are rendered symptom free. Hence, trials using these responder criteria should not be used in trials examining milder symptomatic patients. Moreover, since it is not anticipated that many patients would become symptom free, the entry criteria need to be set somewhere above the minimum required to fulfill these responder criteria.

The choice of the different cut-offs was based on statistical analysis for optimization of the discriminant capacity. The results obtained are close to those expected in the field of OA, i.e. a 20–30% placebo response and a 20–30% treatment effect.[14,15] It is of note that the placebo response in intra-articular studies may be as high as 50%.

The placebo response for hip and knee with an NSAID varied from those obtained in two other subgroups. In the subgroup 'knee-oral specific OA drug', the placebo effect was greater than in the other subgroups. This difference in the placebo effect might be explained by the concomitant therapy. In NSAIDs trials, acetaminophen was the most commonly permitted rescue medication. In the 'specific osteoarthritic drug' trials, both acetaminophen and NSAIDs intake was commonly permitted. Considering the subgroup

'knee-intra-articular specific OA drug', both the placebo and the treatment effect were of greater magnitude than in the other subgroups. The higher placebo effect may be explainable by knee aspiration at the time of the procedure and by a higher placebo effect associated with the route of administration.[17,19]

The higher treatment effect for intra-articular therapy is more difficult to interpret. It is possible that treatment with an intra-articular specific OA drug is of greater magnitude than any drug given orally. However, in the trials reviewed in this effort, the questions concerning the level of pain, functional disability and global assessment were not related to the absolute condition of the patient at the end of the study, but to the relative improvement of the patient; hence, it was not possible to detect worsening during these trials. The authors feel that the sensitivity and specificity obtained in this subgroup should be interpreted with caution.

Specific cut-offs are proposed for the responder criteria. Additional cut-offs were examined, such as a uniform cut-off for all subsets. Unfortunately, these alternative cut-offs showed an important loss of sensitivity and sensitivity.

Because of the lack of available databases, some OA conditions were not evaluated, such as hip oral-specific OA drugs, analgesics, and hand OA.

Two sets of responder criteria were developed in this initiative. The performances of the two sets of criteria in the different evaluated subgroups are similar. One could consider that proposition A is more simple and therefore should be retained. However, the members of the committee, together with other participants in this initiative consider that in some studies changes in functional disability are at least as important as the changes in pain. Hence, proposition B is also included. Finally, these sets of responder criteria should be considered as preliminary. Further studies are needed in order to validate these proposals in other sets of patients and with different drugs.

Acknowledgments

The authors would like to acknowledge the extensive contribution of the members of the OARSI Standing Committee on Clinical Trials, representatives of the health agencies and the representatives of the pharmaceutical industry. This manuscript has been reviewed and approved by the Board of Directors of the OARSI. This effort was supported by educational grants to the OARSI from the pharmaceutical industry participants. The contributors are not listed in order to maintain the anonymity of the products evaluated in this study.

References

1. Altman R, Brandt T, Hochberg M, Moskowitz R. Design and conduct of clinical trials of patients with osteoarthritis. Recommendations from a task force of the Osteoarthritis Research Society. Osteoarthritis Cart 1996;4:217–43.
2. Bellamy N, Kirwan J, Altman R, Boers M, Brandt K, Brooks P, et al. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip and hand osteoarthritis. Results of consensus development at OMERACT III. J Rheumatol 1997;24:799–802.
3. Lequesne M, Brandt K, Bellamy N, Moskowitz R, Menkes CJ, Pelletier JP, et al. Guidelines for testing slow-acting drugs in osteoarthritis (published erratum appears in J Rheumatol Suppl 1994 Dec 21(12): 2395. J Rheumatol 1994;41:65–73.
4. Group of the Respect of Ethics and Excellence in Science (GREES): osteoarthritis section. Recommendations for the registration of drugs used in the treatment of osteoarthritis. Ann Rheum Dis 1996;55:552–7.
5. Lequesne M, Mery C, Samson M, Gerard P. Indexes of severity for osteoarthritis of the hip and knee. Scand J Rheumatol 1997;65:85–9.
6. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt WJ. Validation of WOMAC: A health status instrument for measuring clinically important patient relevant outcomes to anti-rheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1995;15:1833–40.
7. Meenan RF, Mason JH, Anderson JJ, Guccione AA, Kazis IE. AIMS2. The content and properties of a revised and expanded arthritis impact. Measurement Scales Health Status Questionnaire. Arthritis Rheum 1992;35:1–10.
8. Felson DT, Anderson JJ, Boers M, Bombardier C, Furst D, Goldsmith C, et al. American College of Rheumatology preliminary definition of improvement in rheumatoid arthritis. Arthritis Rheum 1995;38:727–35.
9. Van Gestel AM, Prevoo MLL, van't Hof MA, van Rijwuk MH, van de Putte LBA, van Riel PLCM. Development and validation of the European League Against Rheumatism response criteria for rheumatoid arthritis. Comparison with the Preliminary American College of Rheumatology and the World Health Organization/International League Against Rheumatism Criteria. Arthritis Rheum 1996;9:34–40.
10. Bellamy N. Weighting and Aggregation. In: Bellamy N, Ed. Musculoskeletal Clinical Metrology. Boston: Kluwer Academic Publishers 1993:45–52.
11. www.update-softare.com/ccweb/default.htw. (The Cochrane Collaboration web site).
12. Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. Br Med J 1995;310:452–4.
13. Van der Heijde DM, van't Hof MA, van Riel PL, Theunisse LA, Lubberts EW, van Leeuwen MA, et al. Judging disease activity in clinical practice in rheumatoid arthritis: first step in the development of a disease activity score. Ann Rheum Dis 1990;49: 916–20.
14. Altman R, Asch E, Bloch D, Bole G, Borenstein K, Brandt K, et al. Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Arthritis Rheum 1986;29:1039–49.
15. Bolnot-Delmas D, Buch JP, Zeidler H, Dougados M. To 15-8081 in osteoarthritis of hip and knee: a double-blind placebo-controlled multicentre dose-ranging study on analgesia. Pain 1996;64:99–105.
16. Nguyen M, Dougados M, Berdah L, Amor B. Diacerhein in the treatment of osteoarthritis of the hip. Arthritis Rheum 1994;37:529–36.
17. Dougados M, Nguyen M, Listrat V, Amor B. High molecular weight sodium hyaluronate (hyalectin) in osteoarthritis of the knee: a 1 year placebo-controlled trial. Osteoarthritis Cart 1993;1:97–103.

18. Altman R, Moskowitz R, and the Hyalgan® Study Group. Intra-articular sodium hyaluronate (Hyalgan®) in the treatment of patients with osteoarthritis of the knee: a randomized clinical trial. Hyalgan Study Group. J Rheumatol 1998;25:2203–12.

19. Ravaud P, Moulinier L, Giraudeau B, Ayral X, Guerin C, Noel E, et al. Effects of joint lavage and steroid injection in patients with osteoarthritis of the knee: results of a multicenter, randomized, controlled trial. Arthritis Rheum 1999;42:475–82.

Minimal Perceptible Clinical Improvement with the Western Ontario and McMaster Universities Osteoarthritis Index Questionnaire and Global Assessments in Patients with Osteoarthritis

ELLIOT W. EHRICH, GLENN M. DAVIES, DOUGLAS J. WATSON, JAMES A. BOLOGNESE, BETH C. SEIDENBERG, and NICHOLAS BELLAMY

ABSTRACT. Objective. To determine the minimal perceptible clinical improvement (MPCI) in patients with osteoarthritis (OA) with the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) questionnaire, and patient and investigator global assessment of disease status in randomized clinical trials for treatment of OA.

Methods. Subjects with OA of the knee or hip were randomized to receive either rofecoxib 12.5 or 25 mg once daily, ibuprofen 800 mg 3 times daily, or placebo for 6 weeks. The WOMAC and global assessments were completed at baseline and Weeks 2, 4, and 6. A patient global assessment of response to therapy (0 to 4 scale) was used to "anchor" the WOMAC scores. MPCI was defined as the difference in mean change from baseline in WOMAC (100 mm normalized visual analog scale, VAS) between patients with 0 = "None" global response to therapy and patients with 1 = "Poor" global response to therapy.

Results. MPCI was determined to be 9.7, 9.3, and 10.0 mm for the WOMAC pain, physical function and stiffness subscales, respectively, and 11.1 mm for WOMAC question 1: Pain walking on a flat surface. The MPCI for the investigator was 0.4 with investigator assessment of disease status reported on a 0 to 4 Likert scale. Of note, the estimated MPCI for the WOMAC and investigator globals were similar irrespective of treatment, sex, age, or geographic region.

Conclusion. In this analysis, mean changes of roughly 9 to 12 mm (100 mm normalized VAS) on WOMAC scales were perceptible changes to patients with hip and knee OA. A mean decrease of 0.4 in global disease status (0 to 4 Likert scale) as assessed by the investigator corresponded to the patients' MPCI. Understanding the minimal perceptible differences may permit a better assessment of the clinical relevance of therapeutic interventions in OA. (J Rheumatol 2000;27:2635–41)

Key Indexing Terms:
OSTEOARTHRITIS ROFECOXIB CYCLOOXYGENASE INHIBITORS
WOMAC TREATMENT EFFICACY

Therapeutic drug trials for osteoarthritis (OA) often involve a number of efficacy measures including both patient and investigator global assessments of disease status or activity, and response to therapy. Typically, patient self-assessments will include measures of pain and stiffness in affected joint(s), and physical function or disability. These measures must be valid, reliable, and responsive to clinically meaningful change in order to differentiate between treatments[1]. A commonly used OA-specific health status measure is the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)[2]; results from this questionnaire were studied in the context of determining minimal perceptible clinical improvement (MPCI) in patients with OA.

Two important concepts when using health status measures are the minimal clinically important change (or difference), and the minimal perceptible improvement (or difference) to the patient. The former has received some attention in the literature[3-6]. It is important for sample size calculations when planning clinical trials, and also provides the clinician with a basis for reference when considering the effects of treatment in an individual patient. The minimal perceptible difference or change is also important; it represents the difference or change on the measurement scale associated with the smallest change in health status detectable by the patient. It may be less than, the same as, or possibly even greater than the clinically meaningful difference or change. To fully understand differences or changes in health status measures, it is advantageous

From Merck Research Laboratories, Merck & Co, Inc., Rahway, NJ, USA, and The Department of Medicine, Queensland University, Queensland, Australia.

Supported by a grant from Merck and Co., Inc.

E.W. Ehrich, MD; G.M. Davies, DrPH; D.J. Watson, PhD; J.A. Bolognese, MStat; B.C. Seidenberg, MD, Merck and Co., Inc.; N. Bellamy, MD, University of Queensland.

Address reprint requests to Dr. E.W. Ehrich, Merck Research Labs, RY32-629, Rahway, NJ 07065.

Submitted December 7, 1999 revision accepted May 3, 2000.

to know both the minimal clinically important difference and the minimal perceptible difference.

The minimal clinically important and the minimal detectable difference or change on a health status scale can be evaluated by several methods. These involve the use of distribution based statistical measures (e.g., the standardized effect size), or the use of external "benchmarks" or "anchors" as a frame of reference (e.g., normative data, disease classes, or response on an alternative measure)[3-5,7,8]. The latter "anchoring" approach is useful to value differences or changes relative to a metric that can be easily understood by clinicians and care givers, and was utilized for this analysis.

This analysis was designed to determine the minimal perceptible difference or change on the WOMAC VA 3.1[1] and on an investigator global assessment of disease status measure in a population of patients from OA clinical trials, using the anchoring method. Since the interventions employed in the trials were administered following withdrawal of usual therapy and a period of disease flare, or worsening, virtually all patients stayed the same or improved on both measures with treatment. Therefore, we assessed the minimal perceptible clinical improvement.

MATERIALS AND METHODS

Study population. Data from 2 identical 6 week double blind multicenter randomized, placebo controlled, parallel group clinical trials[9] comparing the COX-2 inhibitor rofecoxib (VIOXX®, MK-0966) to ibuprofen in the treatment of OA of the knee or hip were used in the analysis. One trial was conducted in the United States[9] whereas the other trial was performed in 26 countries (Argentina, Brazil, Chile, Costa Rica, Guatemala, Mexico, Peru, Venezuela, Canada, New Zealand, Australia, Austria, Belgium, Czech Republic, Denmark, France, Germany, Holland, Israel, Italy, Norway, Portugal, Spain, Switzerland, Sweden, United Kingdom)[10]. Results from the 2 trials were consistent and data were combined for the analysis of MPCI to maximize statistical precision.

The trials employed a prerandomization washout period during which current therapy for OA was discontinued; during the washout patients had to experience worsening of disease before being eligible for randomization. Patients who were using nonsteroidal antiinflammatory agents (NSAID) to treat their OA prior to the study were required to have (1) a score < 80 mm at the prestudy washout visit on the first item of the WOMAC pain scale (Pain walking on a flat surface), (2) at least a 15 mm increase in the pain walking score after the washout, and (3) a washout score ≥ 40 mm. In addition each patient also had to have an increase of 1 unit in the investigator global assessment of disease status after washout. Prior users of acetaminophen had to have at both the prestudy visit and the visit following washout (1) a score of ≥ 40 mm for pain walking on a flat surface, (2) investigator global assessment of disease status as "fair," "poor," or "very poor", and (3) a score ≥ 40 mm on the Patient global assessment of disease status. Patients who met all entry criteria were randomized to receive, with placebo, rofecoxib 12.5 mg once daily, rofecoxib 25 mg once daily, or ibuprofen 800 mg three times a day.

Measurements. Patient and investigator assessments of response to therapy were made at baseline and at 2, 4, and 6 weeks. Patients completed the following: (1) WOMAC VA 3.1 pain (5 items), stiffness (2 items), and physical function (17 items) scales (0–100 mm VAS for all items; higher scores indicate worse disease status). (2) Patient global assessment of disease status ("Considering all the ways your arthritis affects you, mark an "x" through the line for how well you are doing"; 0–100 mm VAS; higher scores indicate worse status. (3) Patient response to therapy ("How would you rate your response to the study medication you received for arthritis?"; None: no good at all, ineffective drug; poor: some effect, but unsatisfactory; fair: reasonable effect, but could be better; good: satisfactory effect with occasional episodes of pain or stiffness; excellent: ideal response, virtually pain free).

Investigators completed the following: (1) investigator global assessment of disease status scale ("Make a global assessment of the patient's disease status by marking an "x" in one box below"; 0 = very well, 1 = well, 2 = fair, 3 = poor, 4 = very poor). (2) Investigator assessment of patient response to therapy ("Please rate the therapeutic effect of the study medication using the following scale": None: no response, absence of drug effect; poor: minimal response, unacceptable; fair: definite response, but could be better; good: good response, but less than the best possible anticipated response; excellent: the best possible anticipated response, considering the severity and stage of disease).

Note that the response to therapy globals were not measured at baseline since these questions refer to study therapy.

Analytic methods. Change scores were calculated by subtracting the baseline value (at randomization, post washout) for a given metric from the value at Week 6. The patient and investigator response to therapy measures were not performed at baseline; therefore no change scores were calculated for these variables. Missing observations at Week 6 were imputed using the last observed value prior to Week 6 excluding the baseline value.

An "anchoring" method was used to assess the clinical interpretation of the change from baseline for the WOMAC subscale scores, the pain walking on a flat surface item from the WOMAC, investigator global assessment of disease status, and patient global assessment of disease status based upon responses to the patient response to therapy and the investigator assessment of patient response to therapy. Mean WOMAC subscale change scores, and mean change scores on the investigator global assessment of disease status and the patient global assessment of disease status were calculated for each level of the patient global response to therapy. Cumulative distribution functions for the various measures, except for the categorical investigator global assessment of disease status, were generated for each patient and investigator global response category at the 6 week time point.

MPCI, from the perspective of the patient, was defined as the difference in mean change from baseline in WOMAC and global assessment of disease status scores between patients with no response to therapy (response score of 0: "None, no good at all, ineffective drug") and patients with next higher level of response (score 1: "Poor, some effect, but unsatisfactory"). A one unit difference at the lowest end of the global assessment of response scale was used to define MPCI as it reflects minimum (one unit) and lowest degree of improvement that could be detected. MPCI, from the perspective of the investigator, was defined in an analogous fashion using the corresponding investigator global assessment of response to therapy scale.

Analysis of variance (ANOVA) models were evaluated to assess the relationship between change scores on the WOMAC, the investigator global assessment of disease status, and the patient global assessment of disease status with both the patient and investigator global response to therapy measures. Change scores from the WOMAC scales, the investigator global assessment of disease status, and the patient global assessment of disease status at Week 6 were modeled as a function of either patient response to therapy or investigator assessment of response to therapy, treatment, clinical trial, baseline score, age > 65 years, and sex. Pair-wise interactions of the global response to therapy with geographic region [Canada/USA, Latin America, Europe, and Other (South Africa, Israel, New Zealand, Australia)], treatment (placebo, rofecoxib 12.5 mg, rofecoxib 25 mg, ibuprofen), baseline score (categorized into tertiles), age > 65, and sex were also included in the models and assessed at the $\alpha = 0.05$ level. If interactions were not significant for the majority of the outcomes, they were eliminated from the models. All main effects were retained in the final models regardless of statistical significance. Least-square means and differences in least-square means were generated from the final models for each global response to therapy category.

RESULTS

A total of 1545 patients were enrolled in the 2 trials. Data

from 1501 (97%) to 1531 (99%) patients were used in the analyses depending on the particular measure being evaluated; patients were excluded from analyses for missing data. Baseline characteristics of the patients (all treatment groups combined) are shown in Table 1. The mean age was 62 years. The study population was 77% female and 79% white. The mean duration of OA was 9.3 years, with a range of < 6 months to 57 years. Thirteen, 60, and 26% of patients were of American Rheumatism Association Functional Class I, II, and III, respectively. Twenty-three percent of patients had OA of the hip, and 77% of the knee. Prior to the study, 90% of patients used NSAID for their OA, while 10% used acetaminophen. Baseline values for the various assessment measures (all treatment groups combined) are shown in Table 2. Mean baseline scores on the WOMAC (0-100 VAS) were 74, 65, 66, and 64 for pain walking on a flat surface, and the pain, stiffness, and physical function scales, respectively. Mean baseline score for investigator global assessment of disease status (0–4 Likert scale) was 2.9.

The ANOVA models suggested that only interactions between baseline score (based on tertile categories) and MPCI were consistently significant across the various models. As such, the interaction term for baseline score was included in the model. The interaction was primarily due to differences in global response between patients having very severe WOMAC or disease status scores at baseline and those having less severe scores. The MPCI was consistently associated with a greater change from baseline in patients with the worst (highest) scores at baseline. On the other hand, MPCI for patients with the least severe scores at baseline was generally smaller in magnitude.

Table 1. Baseline characteristics of patients. All protocols and treatment groups combined.

Age, yrs	
Mean (SD)	62.4 (9.6)
Range	32–91
Sex (%)	
Male	350 (22.7)
Female	1195 (77.3)
Race (%)	
White	1217 (78.8)
Black	66 (4.3)
Other	262 (17.0)
Duration of OA, yrs	
Mean (SD)	9.3 (8.0)
Range	0–57
ARA Functional Class (%)	
I	206 (13.3)
II	933 (60.4)
III	406 (26.3)
Study joint (%)	
Hip	355 (23.0)
Knee	1190 (77.0)
Prior OA Medication (%)	
NSAID	1392 (90.1)
Acetaminophen	153 (9.9)

Table 2. Baseline values (after washout) of efficacy endpoints*, all treatment groups combined.

Pain walking on a flat surface score, 100 mm VAS	
Mean (SD)	73.6 (15.0)
Median	75.0
Range	23–100
WOMAC pain scale score, 100 mm VAS	
Mean (SD)	65.0 (17.3)
Median	65.4
Range	16–100
WOMAC stiffness scale score, 100 mm VAS	
Mean (SD)	65.9 (20.9)
Median	68.5
Range	0–100
WOMAC physical function scale score, 100 mm VAS	
Mean (SD)	63.9 (18.6)
Median	66.2
Range	7.2–99.9
Patient global assessment of disease status, 100 mm VAS	
Mean (SD)	69.6 (18.4)
Median	71.0
Range	5–100
Investigator global assessment of disease status, 0–4 Likert	
Mean (SD)	2.9 (0.6)
Median	3
Range	1–4

*Patient and investigator global response to therapy not assessed at baseline.

None of the other interaction terms or pairwise comparisons including geographic region, age, sex, or treatment showed evidence of a consistently significant interaction at the 0.05 level across endpoints. In sporadic instances single comparisons between subgroup levels were significant, which was not unexpected given the large number of comparisons performed. The additional interaction terms were therefore dropped from the models.

Table 3 shows the mean change scores on the efficacy measures by the patient global response to therapy measure at 6 weeks, adjusted for treatment protocol, age, sex, and tertile of baseline score. More negative scores indicate greater improvement for all measures. Patients with global responses to therapy of none, poor, fair, good, and excellent had mean investigator global disease status change scores of 0.13, –0.30, –0.96, –1.61, and –2.10, respectively. WOMAC pain walking on a flat surface change scores for patients with none, poor, fair, good, and excellent global responses to therapy were –3.6, –14.6, –27.8, –46.6, and –59.8, respectively. Similar results were seen with the WOMAC pain scale scores, and with the other measures, although the changes were of lesser magnitude compared to the single item of pain walking on a flat surface. The minimal perceptible clinical improvement (defined as the difference in mean change scores between patients with a "none" response and those with a "poor" response on the global responses to therapy) was 0.43 (on a 0–4 Likert scale) for the investigator global disease status measure, and 11.1 (on a 100 mm VAS) for the WOMAC

Table 3. Mean (SE) change score by patient global response to therapy measure, and difference in change score means between categories of patient global response to therapy at 6 weeks*. Note MCPI.

Measure	Patient Response to Therapy at 6 Weeks, mean change from baseline					Difference Between Categories of Patient Response to Therapy			
	None	Poor	Fair	Good	Excellent	MCPI None to Poor	Poor to Fair	Fair to Good	Good to Excellent
Investigator global disease status	0.13 (0.06)	−0.30 (0.06)	−0.96 (0.04)	−1.61 (0.04)	−2.10 (0.07)	0.43 (0.08)	0.67 (0.06)	0.65 (0.05)	0.49 (0.07)
Patient global disease status	5.0 (1.5)	−6.6 (1.4)	−21.0 (1.0)	−41.4 (0.9)	−55.9 (1.7)	11.7 (2.0)	14.3 (1.7)	20.5 (1.2)	14.4 (1.8)
WOMAC pain walking on flat surface	−3.6 (1.6)	−14.6 (1.0)	−27.8 (1.1)	−46.6 (1.0)	−59.8 (1.9)	11.1 (2.1)	13.1 (1.8)	18.8 (1.4)	13.2 (2.0)
WOMAC pain	2.1 (1.4)	−7.5 (1.3)	−20.3 (1.0)	−37.9 (0.9)	−52.0 (1.6)	9.7 (1.9)	12.8 (1.6)	17.6 (1.2)	14.1 (1.7)
WOMAC physical functioning	3.7 (1.3)	−5.7 (1.3)	−15.0 (1.0)	−33.3 (0.8)	−47.7 (1.6)	9.3 (1.8)	9.3 (1.5)	18.3 (1.1)	14.4 (1.6)
WOMAC stiffness	4.2 (1.6)	−5.8 (1.5)	−15.4 (1.1)	−34.8 (1.0)	−51.3 (1.8)	10.0 (2.1)	9.6 (1.7)	19.3 (1.3)	16.5 (1.9)

*Data are adjusted for treatment, protocol age, sex, and tertile of baseline score. More negative values reflect greater improvement. SE: standard error.

pain walking on a flat surface item. The MPCI was of a similar magnitude (9–12 mm) on the other 100 mm VAS scales.

The results of the analyses of the WOMAC pain walking on a flat surface, by patient global response to therapy, are illustrated in Figure 1. Shown are the cumulative distribution functions of change scores at 6 weeks for the patient global assessment of disease status and the various WOMAC subscale responses, according to categories of patient global response to therapy. The MPCI of roughly 10 mm is seen as the difference between the median change scores for the "none" and the "poor" groups. The difference in change scores between other adjacent levels of patient global response (e.g., "poor" to "fair") is generally similar, except between "fair" and "good," where the difference is slightly higher.

Table 4 shows the mean change scores on the efficacy measures by the investigator global response to therapy measure at 6 weeks, adjusted for treatment protocol, age, sex, and tertile of baseline score. More negative scores indicate greater improvement for all measures. Patients rated by the investigator as having global responses to therapy ratings of none, poor, fair, good, and excellent had mean investigator global disease status scores of 0.51, 0.02, −0.66, −1.51, and −2.23, respectively. WOMAC pain walking scores for patients with none, poor, fair, good, and excellent investigator global response to therapy ratings were −1.1, −13.3, −26.2, −42.6, and −54.0, respectively. Similar results were seen with the WOMAC pain scale scores, and with the other measures, although they were of lesser magnitude compared to the sin-

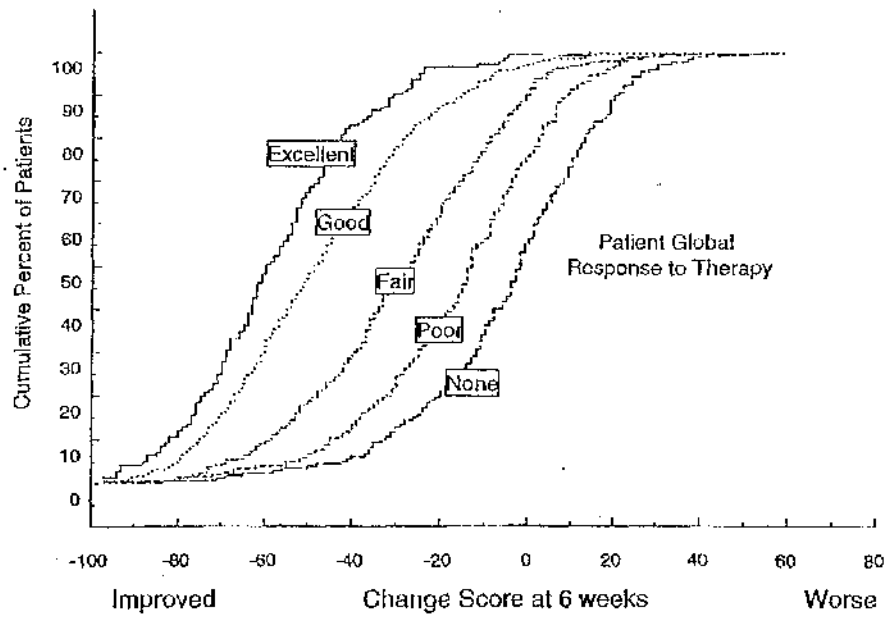


Figure 1. Change scores for WOMAC pain walking on a flat surface, by patient global response to therapy at Week 6.

Table 4. Mean (SE) change score by investigator global response to therapy measure, and difference between change score means between categories of patient global response to therapy at 6 weeks*. Note MCPI.

Measure	Patient Response to Therapy at 6 Weeks, mean change from baseline					Difference Between Categories of Patient Response to Therapy			
	None	Poor	Fair	Good	Excellent	MCPI None to Poor	Poor to Fair	Fair to Good	Good to Excellent
Investigator global disease status	0.51 (0.07)	0.02 (0.07)	−0.66 (0.07)	−1.50 (0.07)	−2.23 (0.07)	0.49 (0.05)	0.68 (0.04)	0.84 (0.03)	0.73 (0.04)
Patient global disease status	6.5 (1.8)	−4.6 (1.4)	−19.0 (1.1)	−37.6 (1.0)	−49.0 (1.3)	11.1 (2.2)	14.4 (1.8)	18.6 (1.4)	11.4 (1.5)
WOMAC pain walking on flat surface	−1.1 (1.9)	−13.3 (1.5)	−26.2 (1.2)	−42.6 (1.0)	−54.0 (1.4)	12.2 (2.4)	12.9 (1.9)	16.4 (1.5)	11.3 (1.6)
WOMAC pain	4.2 (1.6)	−6.6 (1.3)	−18.6 (1.1)	−34.3 (0.9)	−45.5 (1.3)	10.8 (2.1)	12.0 (1.6)	15.7 (1.3)	11.3 (1.4)
WOMAC physical functioning	4.5 (1.6)	−3.2 (1.3)	−14.5 (1.0)	−28.7 (0.9)	−42.3 (1.2)	7.6 (2.0)	11.4 (1.6)	14.2 (1.2)	13.6 (1.4)
WOMAC stiffness	6.3 (1.9)	−4.1 (1.5)	−15.6 (1.2)	−29.9 (1.0)	−43.2 (1.4)	10.4 (2.3)	11.5 (1.9)	14.3 (1.4)	13.3 (1.6)

*Data are adjusted for treatment, protocol age, sex, and tertile of baseline score. More negative values reflect greater improvement. SE: standard error.

gle item of pain walking on a flat surface. In this analysis, the MPCI was 0.49 (on a 0–4 Likert scale) for the investigator global disease status measure. The MPCI on the WOMAC pain walking on a flat surface item was 12.2 (on a 100 mm VAS). The MPCI was of a similar magnitude (7.5–12 mm) on the other 100 mm VAS scales. The results of the analyses of the WOMAC pain walking on a flat surface, by investigator global response to therapy, are shown in Figure 2. The cumulative distribution data show a similar pattern to that seen in Figure 1.

Results of analyses to explore the relationship between outcomes at treatment weeks 2 or 4 versus 6 suggested the MPCI for the WOMAC pain walking on a flat surface item and the investigator global assessment of disease status remained relatively stable across time.

DISCUSSION

Our results suggest that improvements in the range of 8–12 mm on the WOMAC 100 mm normalized VAS and 0.40–0.50 units on the investigator global disease status assessment are considered minimally perceptible to the patient as well as the investigator. To our knowledge, these are the only published data on the MPCI on the WOMAC.

Knowledge of the MPCI on these measures, along with

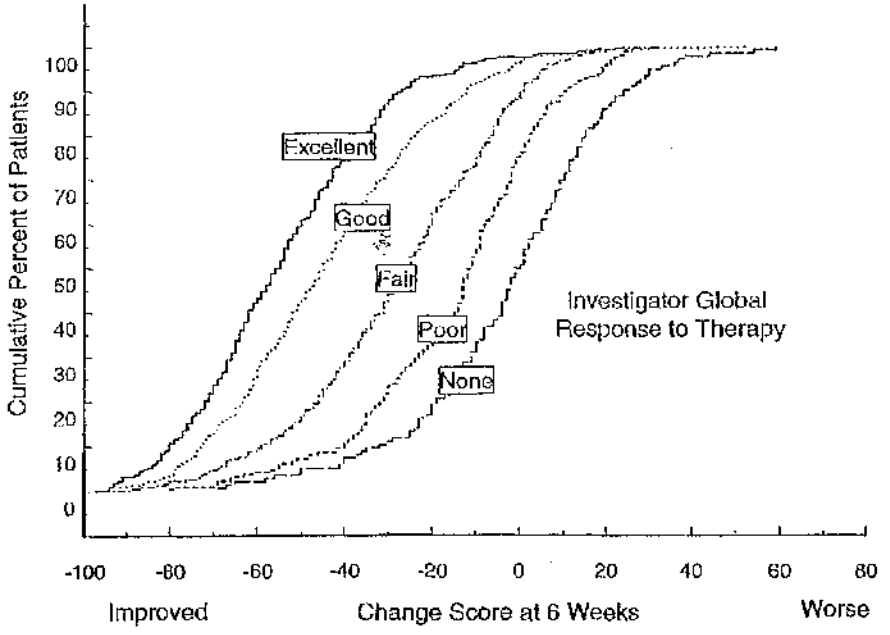


Figure 2. Change scores for WOMAC pain walking on a flat surface, by investigator global response to therapy at Week 6.

knowledge of the minimal clinically important difference (MCID), enables researchers and clinicians to plan and interpret the results of clinical trials that use these assessments. Knowledge of the MCID is necessary for sample size calculations of trials designed to show improved efficacy, and it is also useful to clinicians in interpreting the effect of treatment in an individual patient.

As would be expected, the mean change from baseline observed in the trials used for this analysis exceeded the estimates of MPCI. For example, the mean change from baseline for WOMAC pain walking on a flat surface ranged between 31 and 35 mm for the active treatment groups[9,10] compared with the MPCI estimate of 11 mm. The mean change from baseline with placebo (19 mm), although significantly smaller than that observed with active therapy ($p < 0.001$), also exceeded MPCI. This result suggests the mean placebo response was still perceptible.

Our estimates of minimal *perceptible* improvement are smaller in magnitude than estimates of minimal clinically important difference derived from a consensus development exercise[11]. The consensus exercise found differences of 20 mm, 15 mm, and 0.8 to represent clinically important differences in patient global assessment of disease activity (100 mm VAS), global assessment of pain (100 mm VAS), and investigator global assessment of disease activity, respectively. The difference is not necessarily an inconsistency but rather may highlight that what is minimally perceptible to patients may still be less than a clinically meaningful improvement.

An understanding of MPCI is potentially useful in establishing equivalence criteria for trials intended to show comparability (i.e., no clinically important difference) between active treatments. Such criteria generally specify that the 95% confidence interval of the mean difference in improvement between comparable treatments should not exceed a predefined clinically important difference[12]. Equivalence criteria based on what is minimally perceptible, rather than potentially larger clinically important differences, provide a more conservative or stringent test of equivalence.

Our analysis examined whether there were important differences in results according to geographic region, treatment (including placebo), baseline scores, age, and sex. With the exception of baseline scores, the effects were consistent across these subgroups. A positive correlation between baseline score and subsequent response is consistent with a prior analysis of pain responses in patients with rheumatoid arthritis[13]. The results reported above differ from the finding of Santanello, *et al* that the minimal perceptible patient improvement on an asthma symptom score varied by treatment and by age. In that study, patients treated with placebo and those older than the median age had smaller minimal perceptible improvement values than those undergoing active therapy and younger patients, respectively[14].

Some caution should be used in the interpretation of the "None" categories for the patient and investigator global response to therapy. The scales did not give the option of patients getting "worse"; consequently such patients were pooled with those who had no response to treatment. This could result in the minimally perceptible important change being less than what was indicated by the "anchoring method" used in this analysis. The interaction between baseline scores and MPCI would suggest that there is potentially a threshold level of response to treatment that needs to be attained for a patient to consider himself or herself improved. For example, a 10 mm improvement for a patient with a baseline WOMAC pain score of 90 mm may not be perceptible to the patient, whereas it would be for the patient with a baseline of 60 mm.

The "anchoring method" is one of several methods used to facilitate clinical interpretation of health status measures. Other methods such as collecting normative data, using the standard error of the measurement to assess change, and bench-marking against other diseases have been used to aid in clinical interpretation of health status measures. In this study, we anchored the change on the WOMAC to the patient and investigator assessments of global response to therapy, a method that seems particularly relevant to the interpretation of patient change in therapeutic drug trials.

In summary, this analysis showed that in this population of patients from 2 OA clinical trials, minimal perceptible clinical improvements were observed to be in the range of 8 to 12 units with the WOMAC VAS pain walking on a flat surface item, and the WOMAC pain, stiffness, and physical function scales, and 0.40 to 0.50 units on an investigator global disease status assessment (Likert) scale. With the exception of baseline score, these results did not vary importantly across a variety of patient subgroups. These results provide a reference for the interpretation of changes in these measures in future clinical trials of OA.

ACKNOWLEDGMENT

We acknowledge the advice and review of Drs. Nancy Santanello and Eve Slater. Lisa DeTora provided writing support and edited the document for submission.

REFERENCES

1. Bellamy N. Outcome measurement in osteoarthritis clinical trials. J Rheumatol 1995;22:49-51.
2. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt WL. Validation study of WOMAC: A health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988;15:1833-40.
3. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Control Clin Trials 1989;10:407-15.
4. Lydick E, Epstein RS. Interpretation of quality of life changes. Quality of Life Res 1993;2:221-6.
5. Deyo RA. The significance of treatment effects: The clinical perspective. Med Care 1995;33:286-91.
6. Samsa G, Edelman D, Rothman ML, Williams GR, Lipscomb J, Matchar D. Determining clinically important differences in health status measures. A general approach with illustration to the health utilities index mark II. Pharmacoeconomics 1999;15:141-55.
7. Deyo RA, Inui TS, Leininger J. Physical and psychosocial function

in rheumatoid arthritis: Clinical use of a self-administered health status instrument. Arch Intern Med 1982;142:879-82.
8. Guyatt GH, Jaeschke RJ. Reassessing quality of life instruments in the evaluation of new drugs. Pharmacoeconomics 1997;12:621-6.
9. Saag K, Fischer C, McKay J, et al. MK-0966, a specific Cox-2 inhibitor, has clinical efficacy comparable to ibuprofen in the treatment of knee and hip osteoarthritis in a 6-week controlled clinical trial [abstract]. Arthritis Rheum 1998;41 Suppl: .
10. Day R, Morrison B, Luza A, et al. Rofecoxib, a Cox-2 specific inhibitor, had clinical efficacy comparable to ibuprofen in the treatment of knee and hip osteoarthritis (OA) in a 6-week controlled clinical trial. Arch Intern Med 2000;160:1781-7.
11. Bellamy N, Carette S, Ford PM, et al. Osteoarthritis antirheumatic drug trials. III. Setting the delta for clinical trials — results of a consensus development (delphi) exercise. J Rheumatol 1992;19:451-7.
12. ICH Harmonised Tripartite Guideline. Draft guideline on statistical principles for clinical trials. Federal Register 1997;62:25712-26.
13. Bellamy N, Campbell J, Syrotuik J. Comparative study of self-rating pain scales in rheumatoid arthritis patients. Curr Med Res Opin 1999;15:121-7.
14. Santanello NC, Zhang J, Seidenberg B, Reiss TF, Barber BL. What are minimal important changes for asthma measures in a clinical trial. European Resp 1999;14:23-7.

Towards a Definition of "Difference" in Osteoarthritis

NICHOLAS BELLAMY, ALISON CARR, MAXIME DOUGADOS, BEV SHEA, and GEORGE WELLS

ABSTRACT. To assess existing information regarding detectable differences in osteoarthritis (OA), a systematic literature search was conducted up to December 1999. Thirty-three articles were considered methodologically relevant to the definition and categorization of detectable differences in OA. It was determined that the musculoskeletal literature contains a wealth of information that relates to observed changes, much of which is derived from the clinical trials literature, but there have been relatively few methodological studies that have systematically evaluated the nature, categorization, and relevance of the change. Furthermore, most of those that have been published take the perspective of an individual or groups of experts other than that of the patient. This summary of the current literature reveals that the diverse sources of information go part way towards developing an understanding of detectable differences and their importance in the area of OA research and clinical practice. Stakeholders' interests as well as factors that modulate perceptions of importance need to be taken under consideration. In particular, the patient's perspective of the importance of change at an individual level requires further evaluation. This area of clinical research is relatively underdeveloped, but there is considerable opportunity for progress. (J Rheumatol 2001;28:427–30)

Key Indexing Terms:
DISCRIMINATION OSTEOARTHRITIS
MINIMALLY CLINICALLY IMPORTANT DIFFERENCE

INTRODUCTION

Health status measurement in osteoarthritis (OA) has undergone progressive evolution in the last 60 years[1], with more rapid change in the last 20 years[2]. Core set domains of pain, physical function, patient global assessment, and for studies of one year or longer, imaging, were established by international consensus at the OMERACT 3 conference[3], and subsequently ratified by the Osteoarthritis Research Society International Task Force on clinical trials[4]. The latter were published within guidelines for the execution of future studies, and contained descriptions of relevant measurement techniques. The last 20 years have seen progress in the development of general measures of musculoskeletal status [e.g., the Health Assessment Questionnaire, Arthritis Impact Measurement Scale (AIMS), and AIMS2], generic health related quality of life measures (Medical Outcomes Survey Short Form-36, EUROQOL, NHP, HUI), and disease- specific measures for OA knee and hip disease [Indices of Clinical Severity, Western Ontario and McMaster University OA (WOMAC) Index, WOMBAT Index] and OA hand disease (Algofunctional Index, AUSCAN Index)[2]. Studies of the relative responsiveness of the WOMAC suggest that disease-specific measures may offer advantage over generic arthritis measures and that disease-specific measures are more responsive than generic Health Related Quality of Life (HRQOL) measures[5]. From a conceptual standpoint, the combination of the disease-specific OA measure and a generic HRQOL measure is advantageous in dissecting the impact of interventions on the hierarchy of health states.

There are several approaches to defining detectable and/or important differences in health state. A taxonomy for responsiveness has recently been proposed by Beaton, *et al*[6], which employs a tri-axial classification system according to who is being analyzed (individuals or groups), when the change is being measured (over time/at what point in time), and the type of change being quantified (e.g., observed change versus important change)[6]. The nature of the change being quantified may be considered from various standpoints: (a) minimum change potentially detectable by the instrument; (b) minimum change detectable given the measurement error; (c) observed change in a given population; (d) observed change in those deemed to have improved (estimated change), and/or (e) observed change in those deemed to have an important change[6]. The last 2 types of changes can be viewed from a number of perspectives, including those of the patient, clinician/researcher, payer, and/or society[6].

To assess existing information regarding detectable

From the Department of Medicine, University of Queensland, Queensland, Australia.

N. Bellamy, MD, MSc, FACP, FRCP(Glas,Edin), FRCPC, FAFRM, FRACP, Department of Medicine, University of Queensland; A. Carr, MPH, PhD, University of Nottingham, UK; M. Dougados, MD, Department of Rheumatology, Hôpital Cochin, Paris, France; B. Shea, MSc, Faculty of Medicine, University of Ottawa, Clinical Epidemiology Unit, Loeb Health Research Institute, Ottawa, Canada; G. Wells, PhD, Faculty of Epidemiology and Community Medicine, University of Ottawa, Ottawa, Canada.

Address reprint requests to Prof. N. Bellamy, Department of Medicine, University of Queensland, C Floor, Clinical Sciences Building, Royal Brisbane Hospital, Queensland, Australia.
E-mail: nbellamy@medicine.uq.edu.au

differences in OA, a systematic literature search was conducted in MEDLINE, EMBASE, and Current Contents up to December 1999, using text words for OA and minimal clinically important difference, minimum observable or detectable difference, responsiveness, and improvement criteria. In addition, recent conference proceedings and journals were searched for additional relevant studies. The literature search identified 379 articles. Two independent reviewers assessed the titles and abstracts to determine eligibility. A total of 66 articles were considered potentially relevant and were retrieved for closer examination. Of these 66 articles, 33 were considered methodologically relevant to the definition and categorization of detectable differences in OA. The included articles were then evaluated to determine which concepts of the responsiveness cube were addressed in the publication.

The subsequent review noted that the musculoskeletal literature contains a wealth of information that relates to observed changes, much of which is derived from the clinical trials literature, but there have been relatively few methodologic studies that have systematically evaluated the nature, categorization, and relevance/consequence of the change[2]. Further, most of those that have been published take the perspective of an individual or group of experts other than that of the patient. The articles cited in the following paragraphs are considered relevant to the issue of defining various levels of difference in OA, and are for the most part based wholly or partly on the OMERACT/OARSI core set clinical measures of pain, function, and patient global assessment.

MINIMUM CHANGE POTENTIALLY DETECTABLE BY THE INSTRUMENT

The minimum change potentially detectable (MCPD) is a function of the subscale structure and scale length of the instrument. The smallest detectable difference would be one unit, which in the case of a visual analog scale is 1 mm, and in the case of a Likert scale is equivalent to the smallest numerical difference between adjacent grades defined by the scoring system. In the case of the WOMAC LK 3.1 Index, the scale ranges for the component subscales are as follows: pain 0–20, stiffness 0–8, physical function 0–68, total WOMAC 0–96[5]. Given an MCPD of 1 unit, the minimum percentage change potentially detectable (MPCPD) for the respective elements is as follows: pain 5%, stiffness 12.5%, physical function 1.5%, total WOMAC 1%. By comparison the WOMAC VA3.1 uses scale ranges as follows: pain 0–500, stiffness 0–200, physical function 0–1700, and total WOMAC score 0–2400. The MCPD is 1 mm and the MPCPD values are as follows: pain 0.2%, stiffness 0.5%, physical function 0.06%, total WOMAC score 0.04%. In contrast, the Indices of Clinical Severity[7] are scored on a 0–24 scale, with an option for differences of 0.5 in the physical function component to provide an MCPD of 0.5 and an MPCPD of 2%. It should be noted that the Indices of Clinical Severity are aggregated multidimensional indices and that the total WOMAC score would provide a comparable approach to aggregated measurement, albeit using a different weighting system. With the AUSCAN LK3.0 OA Hand Index[8], the length of the subscales are as follows: pain 0–20, stiffness 0–4, physical function 0–36, total AUSCAN score 0 60. The MCPD is 1 unit and the MPCPD values are as follows: pain 5%, stiffness 25%, physical function 2.8%, and AUSCAN total index score 1.7%. The Algofunctional Index contains 10 questions[9]. The scale range of the Algofunctional Index is 0–30, providing an MCPD of 1 unit and an MPCPD of 3.3%.

MINIMAL CHANGE DETECTABLE GIVEN THE MEASUREMENT ERROR

The measurement error can be subdivided according to several sources including the patient and any independent assessor. Circadian variation in pain and function has been observed in OA of the knee and hand using patient self-report methods and performance based measurement techniques[10,11]. Estimates of measurement error need to consider the volatility of the symptom complex and the way in which variations in a specified time frame might influence the determination[2]. As a result there are relatively few published studies that adequately address this issue.

OBSERVED CHANGE IN A GIVEN POPULATION

There are several sources for observed change in a given population. The majority come from either cohort/observational studies or from published clinical trials. Such studies need to be interpreted in the light of inclusion/exclusion criteria, the nature of the intervention, and the duration of the study. Relatively few clinical trial reports contain an exact description of the method of deriving the minimum clinically important difference sought and which was used in a sample sized calculation[2].

OBSERVED CHANGE IN THOSE DEEMED TO HAVE IMPROVED

The determination of change can be made by the patient, clinician/researcher, payer, or society[6]. It is to be anticipated that the perception of change might be different between different reference groups. In a recently published study evaluating minimum clinically perceptible improvement (MCPI) in OA patients, the MCPI for the WOMAC pain, function, and stiffness subscales (0–100 mm) were 9.7, 9.3, and 10 mm, respectively, while the MPCI for the investigator global assessment of disease status (0–4) was 0.42[12].

OBSERVED CHANGE IN THOSE DEEMED TO HAVE AN IMPORTANT CHANGE

The perceived importance of change may be different for

different stakeholders. In a group of studies published in *The Journal of Rheumatology*[13-15], a 3 round Delphi exercise was used to define minimum clinically important differences (MCID) for clinical trial purposes for a number of outcome measures used in prior OA clinical trials[2]. The median MCID for a comparative study of 2 nonsteroidal antiinflammatory drugs in a double-blind randomized control parallel trial, in the perception of 6 academic rheumatologists experienced in OA clinical trials and based on actual data from 60 patients, were as follows: Doyle Index 5.5, Physicians Overall Assessment of Pain (visual analog scale, VAS) = 15, Physicians Overall Assessment of Pain (Likert Scale, LK) = 0.78, Physicians Overall Assessment of Morning Stiffness (VAS) = 15, Physicians Overall Assessment of Morning Stiffness (LK) = 0.75, Duration of Morning Stiffness (time between arising and improvement in stiffness) = 0.23, Duration of Morning Stiffness (clock time from awaking to when stiffness begins to wear off) = 20, Duration of Morning Stiffness (time between awakening and when patient is limber) = 0.3, Grip Strength (FDA method) = 37.5, Grip Strength (Dictionary of the Rheumatic Diseases Method) = 37.5, Knee Range of Movement = 15, Intercondylar Distance = 6.5, Intermalleolar Distance = 8, Physicians Overall Assessment of Physical Disability (VAS) = 15, Physicians Overall Assessment of Physical Disability (LK) = 0.68, Investigators subject of opinion of Patients General Condition = 0.90, Physicians Estimate of Disease Activity = 0.78, Physicians Global Assessment of Disease Activity (VAS) = 15, Physicians Global Assessment of Disease Activity (LK) = 0.78, Soft Tissue Swelling = 1.50, Patient Pain at Rest (VAS) = 10.5, Patient Pain on Movement (VAS) = 17.5, Patient Overall Assessment of Pain (VAS) = 15, Patient Overall Assessment of Pain (LK) = 0.78, Subjective Pain Evaluation by Patient = 0.78, Patient's Overall Assessment of Morning Stiffness (VAS) = 17.5, Patient's Overall Assessment of Morning Stiffness (LK) = 0.80, Patient's Overall Assessment of Physical Disability (VAS) = 15, Patient's Overall Assessment of Physical Disability (LK) = 0.8, Lequesne Knee Index = 3, Patient Estimate of Disease Activity = 1, Patient's Opinion of General Condition = 0.9, Patient's Global Assessment of Disease Activity (VA) = 20, and Patient Global Assessment of Disease Activity (LK) = 1.

The recent OARSI Response Criteria Initiative (RCI) has permitted the development of response criteria for clinical trials in OA based on an analysis of 14 placebo controlled clinical trials (totaling 1886 patients). The criteria were presented at the OARSI International Conference in Vienna and use a tree format to categorize patients as responders or nonresponders according to 2 sets of class-specific criteria[16]. The first set of responder criteria are based on a high pain response, or alternatively a lower level of response on at least 2 of the 3 domains: pain, function, and patient global assessment. In contrast, the second set of responder criteria are based on a high level of response in pain or function, or alternatively, a lower level of response on at least 2 of the 3 domains: pain, function, and patient global assessment. These 2 different criteria sets accommodate the dynamic profiles of different classes of interventions. In both sets of criteria, a response is defined by a combination of both absolute and percentage change. As a consequence, they are applicable only to those patients whose symptom severity is such that they could qualify as a responder should their condition improve sufficiently. It is anticipated that the OARSI criteria will require further validation using additional data sets. Doubtless there will be further debate regarding the use of absolute and/or percentage change, the implications of incorporating initial and/or final values, and the implications of dichotomization. Nevertheless, the OARSI responder criteria represent an initial attempt to address the complex and challenging problem of dichotomizing continuous variables, in order to define clinically important changes in health status.

An alternative approach is to provide individual clinical profiles of OA patients to key informants and require them to categorize the patients according to whether they, the key informants, regard the change as being clinically important. Such a project was completed immediately prior to OMERACT 5. The study was based on the WOMAC Index and patient global assessments, and employed a 3-round Delphi exercise to facilitate consensus building. A report is pending.

PATIENTS' DEFINITIONS OF CHANGE

Most assessments of treatment efficacy within clinical trials, and to a lesser extent in clinical practice, are based on clinicians' definitions of clinically important change. Little is known about the degree to which clinicians' and patients' perceptions of clinically important change are concordant, but there is evidence from a number of studies that clinicians are poor judges of the degree of pain suffered by patients, their quality of life, and the relative importance of different treatment outcomes[17-21]. Qualitative research with patients with rheumatoid arthritis and their clinicians has highlighted differences in the ways in which patients and clinicians construct and evaluate disease activity, with patients focusing on the personal consequences in terms of pain and functional limitations and clinicians using biological indicators[22]. In OA, the criteria by which patients judge treatment efficacy appear to focus entirely on pain and function and are very specific, for example, being able to sit through one television program in comfort or being able to walk to a particular shop. There is also some suggestion that patients make "allowances" for treatments they particularly want to work, altering their efficacy criteria when the treatment fails their initial evaluation. These data suggest that patients' assessments of clinically important differences are highly individualized and inconsistent across different inter-

ventions. Quantifying patients' minimum clinically important differences to interpret the results of clinical trials, or as the basis of sample size calculations, may therefore be more complex than using differences derived mathematically from outcome measures or from groups of clinicians. Nevertheless, they are important predictors of health service use. Demand for medical care and treatment change is driven by patients' perceptions of treatment efficacy, and some attempt should be made to include them, particularly in clinical practice.

Issues surrounding the determination of the clinical importance and consequence of structural conservation have received little attention. It will be important to develop outcome measurement strategies for longterm studies. The issues are subtly different in situations where the progression of structural damage may be prevented, slowed, arrested, or reversed. Traditional measures of pain, patient global assessment and especially physical function will be relevant. So too may be the propensity for interventions to reduce the need for total hip replacement surgery[24], although the timing of this endpoint, while clinically relevant, is potentially subject to effects that relate more closely to the health care system in which treatment is being delivered than to actual health status of the individuals concerned.

These diverse sources of information go part way towards developing an understanding of detectable differences and their importance in the area of OA research and clinical practice. Stakeholder interests as well as factors that modulate perceptions of importance need to be taken into consideration. In particular, the patient's perspective of the importance of change at an individual level requires further evaluation. This area of clinical research is relatively underdeveloped, but there is considerable opportunity for progress.

REFERENCES

1. Bellamy N, Buchanan WW. Dr. Douglas Taylor and half a century of musculoskeletal Kelvinism. J Clin Rheumatol 1988;7:272-7.
2. Bellamy N. Musculoskeletal clinical metrology. Lancaster: Kluwer Academic Publishers; 1993.
3. Bellamy N, Kirwan J, Boers M, et al. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip and hand osteoarthritis. Consensus development at OMERACT III. J Rheumatol 1997;24;799-802.
4. Osteoarthritis Research Society (OARS) Task Force Report: Design and conduct of clinical trials in patients with osteoarthritis; recommendations from a task force of the Osteoarthritis Research Society. Osteoarthritis Cartilage 1996;4:217-43.
5. Bellamy N. WOMAC user guide IV. Brisbane, Australia: 2000;43.
6. Beaton DE, Bombardier C, Katz J et al. Looking for important change/difference in studies of responsiveness. J Rheumatol 2001;28:400-5.
7. Lequesne MG, Mery C, Samson M, et al. Indexes of severity for osteoarthritis of the hip and knee. Validation — value of comparison with other assessment tests. Scand J Rheumatol 1987;65 Suppl:85-9.
8. Bellamy N, Campbell J, Haraoui B, et al. Development of the Australian/Canadian (AUSCAN) Osteoarthritis Hand Index [abstract]. Arthritis Rheum 1997;40 Suppl:S110.
9. Dreiser RL, Maheu E, Guillou GB, et al. Validation of an algofunctional index for osteoarthritis of the hand. Rev Rheum (English ed.) 1995;62 Suppl 1:43S-54S.
10. Bellamy N, Sothern RB, Campbell J. Rhythmic variations in pain perception in osteoarthritis of the knee. J Rheumatol 1990; 17:364-72.
11. Bellamy N, Campbell J, Sothern R, et al. Circadian rhythm in pain, stiffness and manual dexterity in osteoarthritis of the hand: relationship between discomfort and disability. Chronobiology International 1997;24 Suppl 1:16.
12. Davies G, Ehrich E, Watson D, et al. Minimum perceptible clinical improvement in osteoarthritis: the WOMAC Questionnaire and Global Assessments [abstract]. Ann Rheum Dis 1999;P1078:255.
13. Bellamy N, Carette S, Ford PM, et al. Osteoarthritis antirheumatic drug trials. I. Effects of standardisation procedures on observer dependent outcome measures. J Rheumatol 1992;19:436-3.
14. Bellamy N, Carette S, Ford PM, et al. Osteoarthritis antirheumatic drug trials. II. Tables for calculating sample size for clinical trials. J Rheumatol 1992;19:444-50.
15. Bellamy N, Carette S, Ford PM, et al. Osteoarthritis antirheumatic drug trials. III. Setting the delta for clinical trials — results of a consensus development (delphi exercise). J Rheumatol 1992;19:451-7.
16. Dougados M, Leclaire P, van der Heijde D, et al. Response criteria for clinical trials on osteoarthritis of the knee and hip. Osteoarthritis Cart 2000 (In Press).
17. Donovan J. Patient education and the consultation: the importance of lay beliefs. Ann Rheum Dis 1991;50:418-21.
18. Kwoh K, Feinstein A, Gifford R, et al. Physician-patient agreement on patients' priorities for important treatment outcomes [abstract]. Arthritis Rheum 1987;30 Suppl 4:S101.
19. Pearlman RA, Uhlmann RF. Quality of life in chronic diseases: perceptions of elderly patients. J Gerontol 1988;43:M25-30.
20. Daltroy L, Eaton H, Hashimoto H, Liang MH. Rheumatoid arthritis patients' expectations for treatment outcomes [abstract]. Arthritis Rheum 1995;38 Suppl:S310.
21. Woolhead GM, Carr AJ, Wilkinson M, Bannister GC, Ackroyd CE. Expectations of treatment and outcome in patients awaiting joint replacement surgery and general orthopaedic outpatients [abstract]. Arthritis Rheum 1996;39 Suppl:S174.
22. Mitchell H, Stowers K, Hughes RA, Carr AJ. Patients' self-assessment of disease activity and severity in rheumatoid arthritis [abstract]. Arthritis Rheum 1999;42 Suppl:S274.
23. Carr AJ, Hughes RA, Stowers K. What influences placebo response to a complementary therapy, glucosamine sulfate? [abstract]. Arthritis Rheum 2000;43 Suppl:S223.
24. Dougados M, Gueguen A, Nguyen M, et al. Requirement for total hip arthroplasty: an outcome measure of hip osteoarthritis. J Rheumatol 1999;26:855-61.

Minimal Clinically Important Difference Module: Summary, Recommendations, and Research Agenda

INTRODUCTION

In preparation for the conference, methodological papers on minimal clinically important differences (MCID) and the current status and need for MCID on core measures in the 4 content areas of interest [osteoarthritis (OA), rheumatoid arthritis (RA), osteoporosis (OP), and low back pain (LBP)] were written and distributed to the conference participants. At the conference, these papers were presented in plenary and the participants were then divided into 8 breakout groups (3 groups each for OA and RA, and one group each for OP and LBP) to further consider the issues. Questionnaires were distributed to participants in the breakout sessions to help focus discussions during these sessions. All questionnaires had a similar core set of questions on MCID, as well as questions that were specific to the content area (OA, RA, OP, LBP). Completed questionnaires were collected by the breakout session leaders at the end of the sessions and the information was entered into a database and analyzed. At the closing plenary session of the conference, the results of the breakout questionnaire and discussions were presented and conference participants voted on key issues associated with MCID.

This paper presents results from the questionnaire and voting, summarizes the associated discussions, and identifies some areas for further research.

RESPONSES TO BREAKOUT QUESTIONNAIRE

A total of 136 questionnaires were returned. Most were from the larger number of RA and OA sessions (RA 58; OA 54; OP 19; LBP 5). The specific results presented here primarily relate to RA and OP.

Of the 30 cells in the "cube" defined by who is the focus (groups, individuals), which scores are contrasted (differences between, changes within, both), and what type of change (minimum potentially detectable, minimum actually detectable beyond error, observed in the population, observed in those estimated to differ/change, observed in those estimated to have important difference/change), the cell of most interest was the "individual" setting for "within change" scores on "important change."

In the RA breakout sessions, participants discussed the classification in the cube of the current RA response criteria according to type of change/difference. The response to the question, "Where are the RA criteria currently placed in the cube," indicated that the majority of participants believed that the ACR20 improvement criteria and the EULAR response criteria were considering change/difference observed in those estimated to have changed or estimated to have an important change (ACR20 52%; EULAR 73%). Based on the assumption that ACR and EULAR criteria have defined response or improvement corresponding to MCID, the participants ranked priority areas for further research. The percentage of priority rankings (rank 1 or 2) for the different areas considered were as follows: defining major improvement (25%); studies that focus on defining individual response as opposed to group change or difference (18%); studies that attempt to define MCID for individual elements of the core set including functional status measures (16%); further validation of ACR/EULAR definitions against independent definitions of response (15%); and studies that focus on whether thresholds for response differ for different core set items (14%). Only 1% gave a priority ranking to studies that evaluate whether core set items have particular statistically measurable thresholds for MCID. An area of study not listed but given a priority ranking by 5% under category of other was validating short term response criteria in predicting longterm outcome received 5% priority ranking.

The issue of major improvement was further explored, with 76% indicating that it was important or useful to establish criterion for a major clinically important improvement as well as a minimal clinically important difference. A qualitative analysis of participants' comments in considering major improvement indicated the following supporting themes: MCID is only a lower bound of improvement change; major change comes after determination and understanding of MCID; treatment decisions are more often made based on major change; and major change is important in interpreting trials of 2 active treatments.

In the OA breakout sessions, the adequacy of current estimates for the OA core set of measures (pain, function, and patient global assessment) were considered from the perspective of the different types of change, namely: (1) minimum potentially detectable; (2) minimum actually detectable beyond error; (3) observed in the population; (4) observed in those estimated to differ/change; and (5) observed in those estimated to have important difference/change. The current estimates for pain were considered at least adequate ("very adequate or adequate") by over 70% of the respondents for type 1, 2, and 3 change.

Only 42% and 23% considered it at least adequate for the categories observed in those estimated to differ/change and observed in those estimated to have important difference/change, respectively. For function, about 70% found it at least adequate for all types of change except for the category observed in those estimated to have important difference/change, which received only 22%. Patient global assessment had a similar pattern to pain with a large percentage indicating it was at least adequate for type 1, 2, and 3 change (over 65%) but a smaller percentage for type 4 (34%) and type 5 (21%) change.

The small number of participants in the OP and LBP breakout groups made it difficult to analyze and interpret their individual breakout session results. The information obtained in these sessions will be relayed to and considered by their respective societies and interest groups as a basis for possible further studies.

In all the breakout sessions, the participants were asked whether a MCID should be defined in terms of percentage change only, absolute change only, or both. The vast majority indicated both (85%), with an equal percentage of respondents indicating support for percent (7.5%) or absolute (7.5%) change only.

RESPONSES TO PLENARY QUESTIONS

Four questions were posed and voted on in the final plenary sessions. The questions were designed to confirm discussions that took place among the participants and the interpretation of the questionnaire results. The goal was to set a broad overview on a research agenda.

Question 1: Do you support the development of clinical response criteria for individuals in other diseases?

Yes	96%
No	1%
Don't know	3%

The MCID module concentrated on 4 content areas (OA, RA, OP, LBP). To address whether other areas should consider clinical response criteria in this way, a question was posed to draw on the various expertise of the participants, as well as the information they were provided with and their specific experience at the OMERACT conference. The vote by all the conference participants at the final plenary session resulted in 96% supporting the development of clinical response criteria for individuals in other diseases.

Question 2: Do you agree that it is important to define "major" clinical important improvement for RA?

Strongly agree	47%
Agree	33%
Neutral	13%
Disagree	3%
Strongly disagree	2%
Don't know	1%

The results of the questionnaire for the RA breakout sessions indicated that "major" clinically important improvement may be an important area of consideration for research. This question was posed to and voted on by all the conference participants at the final plenary session, with a resulting 80% in agreement.

Question 3: Do you agree that it is important to validate short term response/improvement criteria in predicting longterm outcome?

Strongly agree	61%
Agree	23%
Neutral	8%
Disagree	2%
Strongly disagree	4%
Don't know	2%

The issue of validating short term response/improvement criteria in predicting longterm outcome received an important priority ranking but was essentially considered by only one breakout session. This question was posed to and voted on by all the conference participants, with a resulting 84% in agreement.

Question 4: In OA should any response criteria developed be defined in terms of...?

Percent change alone	5%
Absolute change alone	4%
Both	91%

Based on the breakout questionnaire, a large majority of participants indicated that a MCID should be defined in terms of both percentage change and absolute change only. This is an important concept in the development and interpretation of MCID and confirmation of this finding was sought for OA response criteria. This question was posed to and voted on by all the conference participants, with a resulting 91% indicating that both absolute and relative should be considered. After the vote, Maxime Dougados presented the recent work and decisions made by OARSI in which both percentage and absolute change were used in the definition of OA response criteria.

RESEARCH AGENDA OVERVIEW:

1. Develop clinical response criteria for individuals in other diseases?
2. Consider both relative and absolute change in developing response criteria.
3. Consider "major" clinically important change in the further development of a clinical response criteria.
4. Validate short term response criteria in predicting longterm outcome?
5. Consider the patient perspective in developing response criteria.

CONCLUSION

Progress has been made in considering changes/differences related to clinical outcomes of interest in some of the disease areas. Through a multidisciplinary approach at OMERACT involving academic investigators, clinicians,

and regulatory experts, it is anticipated that this work will progress further in areas in which it is more established, with the possible development of "major" response criteria, and be initiated in areas in which it needs more consideration. During the discussions, two important themes evolved that were in need of more consideration — taking a patient perspective of response and validating the longterm clinical consequences of short term response criteria.

GEORGE WELLS, PhD, Department of Epidemiology and Community Medicine, University of Ottawa, Clinical Epidemiology Unit, Loeb Health Research Institute; **JENNIFER ANDERSON,** PhD, The Arthritis Center, School of Medicine, Boston University, Boston, MA, USA; **DORCAS BEATON,** Institute for Work and Health, Department of Occupational Therapy, University of Toronto, Toronto, Canada; **NICHOLAS BELLAMY,** MD, MSc, FACP, FRCP(Glas, Edin), Department of Medicine, University of Queensland, Queensland, Australia; **MAARTEN BOERS,** MSc, MD, PhD, Vrije Universiteit Amsterdam, The Netherlands; **CLAIRE BOMBARDIER,** MD, FRCPC, Institute for Work and Health, Institute of Medical Sciences, University of Toronto, Toronto, Canada; **FERDINAND BREEDVELD,** Department of Rheumatology, University Hospital, Leiden, The Netherlands; **ALISON CARR,** PhD, Kings College Hospital, Dulwich, London, UK; **ANN CRANNEY,** MSc, MD, FRCPC, Clinical Epidemiology Unit, Loeb Health Research Institute, Ottawa, Canada; **MAXIME DOUGADOS,** MD, Department of Rheumatology, Hôpital Cochin, Paris, France; **DAVID FELSON,** MD, The Arthritis Center, School of Medicine, Boston University, Boston, MA, USA; **JOHN KIRWAN,** MD, Rheumatology Unit, University of Bristol, Division of Medicine, Bristol Infirmary, Bristol, UK; **MICHAEL SCHIFF,** MD, Denver Arthritis Clinic, Denver, CO, USA; **BEVERLEY SHEA,** MSc, RN, Clinical Epidemiology Unit, Loeb Health Research Institute, Faculty of Medicine, Ottawa Hospital, Ottawa, Canada; **LEE SIMON,** MD, Beth Israel Deaconess Medical Center, Boston, MA, USA; **JOSEF SMOLEN,** MD, Department of Medicine, Krankenhaus Lainz, Vienna, Austria; **VIBEKE STRAND,** MD, Department of Medicine, Stanford University, Stanford, CA, USA; **PETER TUGWELL,** MD, MSc, FRCPC, Clinical Epidemiology Unit, Loeb Health Research Institute, Faculty of Medicine, Ottawa Hospital; **PIET van RIEL,** MD, Department of Rheumatology, University Hospital Nijimegen, The Netherlands; **VIVIAN A. WELCH,** MSc, Clinical Epidemiology Unit, Loeb Health Research Institute.

Address reprint requests to Dr. G. Wells, Department of Epidemiology and Community Medicine, University of Ottawa, Health Sciences Complex, 451 Smyth Road, Ottawa, Canada. E-mail: gwells@uottawa.ca

OsteoArthritis and Cartilage (2004) 12, 389–399

doi:10.1016/j.joca.2004.02.001

Osteoarthritis and Cartilage

International Cartilage Repair Society

OSTEOARTHRITIS

OMERACT-OARSI Initiative: Osteoarthritis Research Society International set of responder criteria for osteoarthritis clinical trials revisited.

T. Pham MD†, D. van der Heijde MD, PhD‡, R. D. Altman MD§, J. J. Anderson MD¶, N. Bellamy MD, MSc, FACP, FRCP, FRCPC, FAFRM, FRACP‖, M. Hochberg MD, MPH††, L. Simon MD‡‡, V. Strand MD§§, T. Woodworth¶¶ and M. Dougados MD‖‖*

† *René Descartes University, Cochin Hospital, Paris, Aix-Marseille II University, Conception Hospital, AP-HP, Marseille, France*

‡ *Department of Rheumatology, University Hospital, Maastricht, The Netherlands*

§ *University of Miami, Department of Veterans Affairs Medical Center, Miami, FL, USA*

¶ *Clinical Epidemiology Research and Training Unit, Boston Medical Center, Boston, MA, USA*

‖ *Department of Medicine, University of Queensland, Royal Brisbane Hospital, Queensland, Australia*

†† *Division of Rheumatology and Clinical Immunology, Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, USA*

‡‡ *Medicine Department, Beth Israel Deaconess Medical Center, Boston, MA, USA*

§§ *Biopharmaceutical Consultant, Portola Valley, CA, USA*

¶¶ *Pfizer Global Research and Development, New London, CT, USA*

‖‖ *René Descartes University, Cochin Hospital, Paris, France*[1]

Summary

Background: The OARSI Standing Committee for Clinical Trials Response Criteria Initiative had developed two sets of responder criteria to present the results of changes after treatment in three symptomatic domains (pain, function, and patient's global assessment) as a single variable for clinical trials (1). For each domain, a response was defined by both a relative and an absolute change, with different cut-offs with regard to the drug, the route of administration and the OA localization.

Objective: To propose a simplified set of responder criteria with a similar cut-off, whatever the drug, the route or the OA localization.

Methods: Data driven approach:

(1) Two databases were considered

- The 'elaboration' database with which the formal OARSI sets of responder criteria were elaborated and
- The 'revisit' database.

(2) Six different scenarios were evaluated:

- The two formal OARSI sets of criteria
- Four proposed scenarios of simplified sets of criteria

Data from clinical randomized blinded placebo controlled trials were used to evaluate the performances of the two formal scenarios with two different databases ('elaboration' *versus* 'revisit') and those of the four proposed simplified scenarios within the 'revisit' database. The placebo effect, active effect, treatment effect, and the required sample arm size to obtain the placebo effect and the active treatment effect observed were the performances evaluated for each of the six scenarios. *Experts' opinion approach:* Results were discussed among the participants of the OMERACT VI meeting, who voted to select the definite OMERACT-OARSI set of criteria (one of the six evaluated scenarios).

*For correspondence and request for reprint: Professor Maxime Dougados, Rheumatology B Department, Cochin Hospital, 27 rue du faubourg Saint Jacques, 75014 Paris, France. Tel.: +33-1-58-41-25-62; Fax: +33-1-43-54-92-56; E-mail: maxime.dougados@cch.ap-hop-paris.fr

[1] Name of department and institution to which the work should be attributed: René Descartes University, AP-HP, Cochin Hospital, Rheumatology B Department, 27 rue du faubourg Saint Jacques, 75014 Paris, France.

Received 17 September 2002; revision accepted 3 February 2003.

Results: Data driven approach: Fourteen trials totaling 1886 OA patients and fifteen studies involving 8164 OA patients were evaluated in the 'elaboration' and the 'revisit' databases respectively.

The variability of the performances observed in the 'revisit' database when using the different simplified scenarios was similar to that observed between the two databases ('elaboration' *versus* 'revisit') when using the formal scenarios. The treatment effect and the required sample arm size were similar for each set of criteria. *Experts' opinion approach:* According to the experts, these two previous performances were the most important of an optimal set of responder criteria. They chose the set of criteria considering both pain and function as evaluation domain and requiring an absolute change and a relative change from baseline to define a response, with similar cut-offs whatever the drug, the route of administration or the OA localization.

Conclusion: This data driven and experts' opinion approach is the basis for proposing an optimal simplified set of responder criteria for OA clinical trials. Other studies, using other sets of OA patients, are required in order to further validate this proposed OMERACT – OARSI set of criteria.

Key words: Osteoarthritis, Outcomes, Clinical Trials Response Criteria Initiative.

Introduction

The Osteoarthritis Research Society International (OARSI) Standing Committee for Clinical Trials Response Criteria Initiative and the Outcome Measures in Rheumatology (OMERACT) committee, in concert with the international rheumatology community, has led to the development of a uniform core set of outcome measures for osteoarthritis (OA)[1–4]. One of the objectives was to propose a set of criteria for measurement based on multiple domains to present the results of changes after treatment in symptomatic parameters as a single variable for clinical trials. The symptomatic variables selected by both the OMERACT and OARSI societies were: pain, functional impairment and patient's global assessment.

Based on data from clinical trials, two sets of responder criteria (formal OARSI criteria) that can categorize an individual's response to treatment in a clinical trial have been developed[5] (Fig. 1).

The main characteristics of the proposed sets of criteria were the following:

- They covered three domains: pain, function and patient's global assessment.
- For each of these domains, a response was defined by both a relative and an absolute change.
- The cut-offs that defined a relevant change differed with regard to:
  - OA localization (*e.g.* hip *vs* knee),
  - evaluated study drug (*e.g.* NSAIDs *vs* specific anti-OA drug),
  - route of administration (*e.g.* per os *vs* intra-articular),
  - specific domain (pain, function, patient's global assessment).

The choice of the different cut-offs for the formal OARSI set of criteria was based on statistical analysis for optimization of the discriminant capacity. The preliminary attempts at uniform cut-off of all subsets showed a lesser placebo and active treatment effect of the set of criteria considered relevant by the members of the steering committee.

The main objective of this study was to evaluate the performances of the two previous formal OARSI sets of criteria and the performances of the modified ones, proposed by the scientific OMERACT committee. The aim of the proposed modifications was to simplify the presentation of the set of criteria, by evaluating different scenarios whatever the OA localization, whatever the evaluated drug, whatever the route of administration, and with similar cut-offs for the different domains.

Methods

PROPOSED SET OF CRITERIA

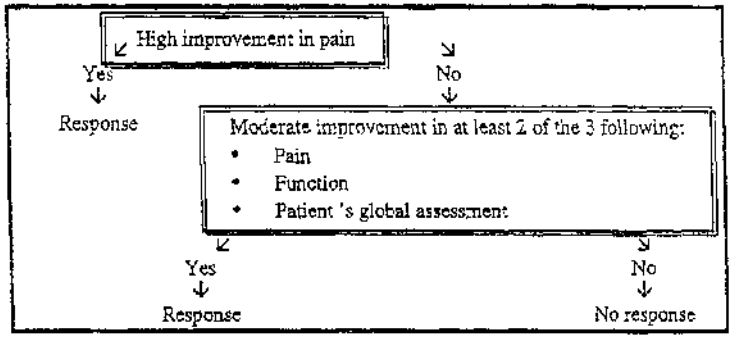
Six different scenarios were evaluated. The first two scenarios were the two propositions (A and B) of the formal OARSI set of criteria[5] (Fig. 1). The four other scenarios (scenarios C to F) were proposed by the OMERACT scientific committee. Their main characteristic was that they used a uniform cut-off whatever the OA localization, whatever the study drug and whatever the route of administration, unlike the formal OARSI set of criteria (Fig. 2). Scenarios A, C and E considered pain at the first responder step (high improvement), and scenarios B, D and F considered pain or function (Fig. 3). Scenarios C and D, as the formal OARSI set of criteria did, considered relative change (percentage of change during the study) and absolute change (absolute change during the study) in the variable to define a response, whereas scenarios E and F considered only relative change to define such response.

The study approach was both data driven and used an experts' opinion approach.

DATA DRIVEN APPROACH

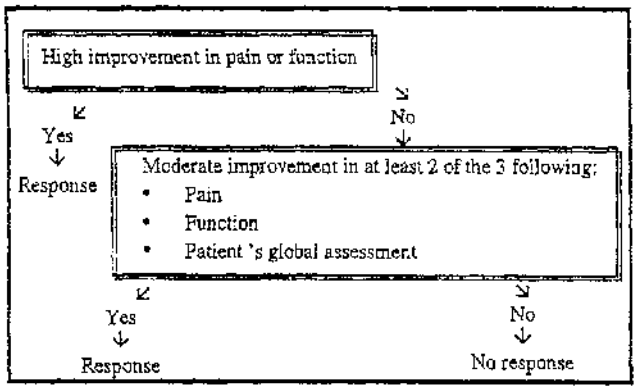
Two databases from clinical randomized placebo controlled trials were used:

- The initial one used to elaborate the formal set of criteria, known here as the 'elaboration database'.
- The second one is labeled the 'revisit database'. Drug companies who had conducted positive randomized placebo controlled trials in OA of a minimum 4-week duration were invited to revisit their database. The definition of 'positive' was based on a p value <0.05 for the *a priori* chosen primary criterion of the trial. Only the intention-to-treat analysis trials using the Last Observation Carried Forward technique were used. The participating drug companies were invited to provide anonymous information: OA localization, route of administration, characteristics of the study drug (analgesic, NSAID, Specific OA drug), study duration, number of patients in the placebo group and in the active treatment group, tools used to evaluate pain (*e.g.* pain VAS, Likert scale, WOMAC pain subscale), function (*e.g.* WOMAC function subscale) and patient's global assessment (*e.g.* VAS, Likert scale)[6], and time of collection of these different tools. Because of confidentiality, no demographic data, such as age, gender, body mass index, nor baseline values were asked to the drug companies. The drug was not identified by name, but only by class of agent (*e.g.* NSAIDs,



Subgroup	*High improvement in pain*				*Moderate improvement in* *Pain*		*Function*	
	Relative change*	Absolute change**	Relative change	Absolute change	Relative change	Absolute change	Relative change	Absolute change
Knee, oral NSAIDs	45	20	15	10	30	15	35	10
Knee, oral specific drug	55	30	35	10	15	20	15	15
The 3 above groups together	55	30	35	15	15	20	15	15
Knee, intra-articular specific drug	40	30	35	15	35	10	30	10

Proposition A Optimal cut-offs to be applied for the OARSI Responder Criteria



Subgroup	*High improvement in* *Pain*		*Function*		*Moderate improvement in* *Pain*		*Function*		*Global assessment*	
	Relative change*	Absolute change**	Relative change	Absolute change	Relative change	Absolute change	Relative change	Absolute change	Relative change	Absolute change
Hip, NSAIDs	50	30	50	20	25	15	20	10	20	10
Knee, oral NSAIDs	50	20	60	20	30	15	20	20	25	10
Knee, oral specific drug	55	30	50	20	30	20	20	20	20	15
The 3 above groups together	55	30	50	20	30	15	20	20	20	15
Knee, intra-articular specific drug	50	30	60	20	20	20	30	10	30	10

Proposition B Optimal cut-offs to be applied for the OARSI Responder Criteria

* Relative change: percentage of change during the study (final minus baseline over baseline × 100)

** Absolute change: absolute change during the study (final minus baseline on a 0-100 interval scale)

Fig. 1. OARSI Formal set of criteria: Scenarios A and B.

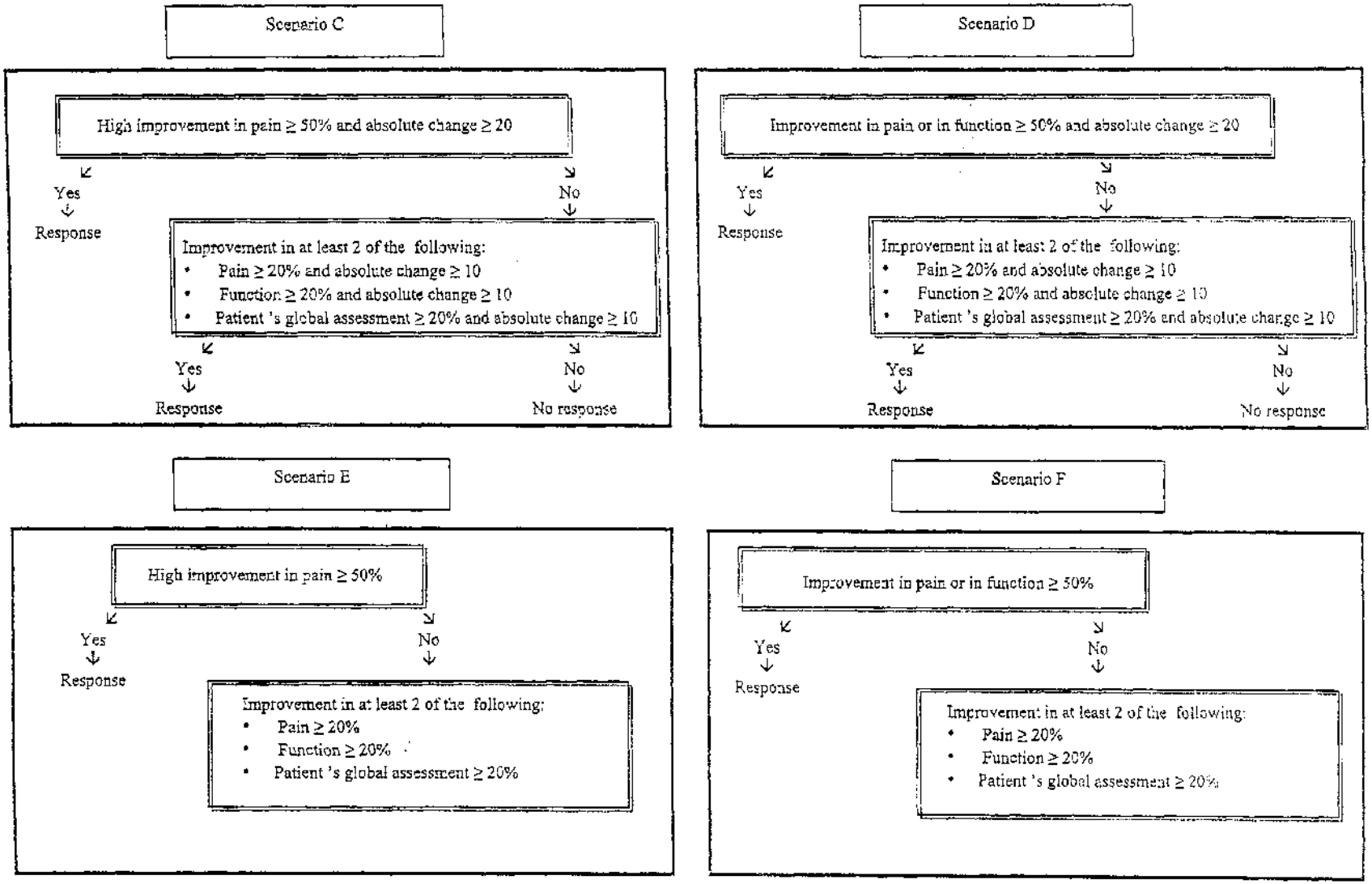
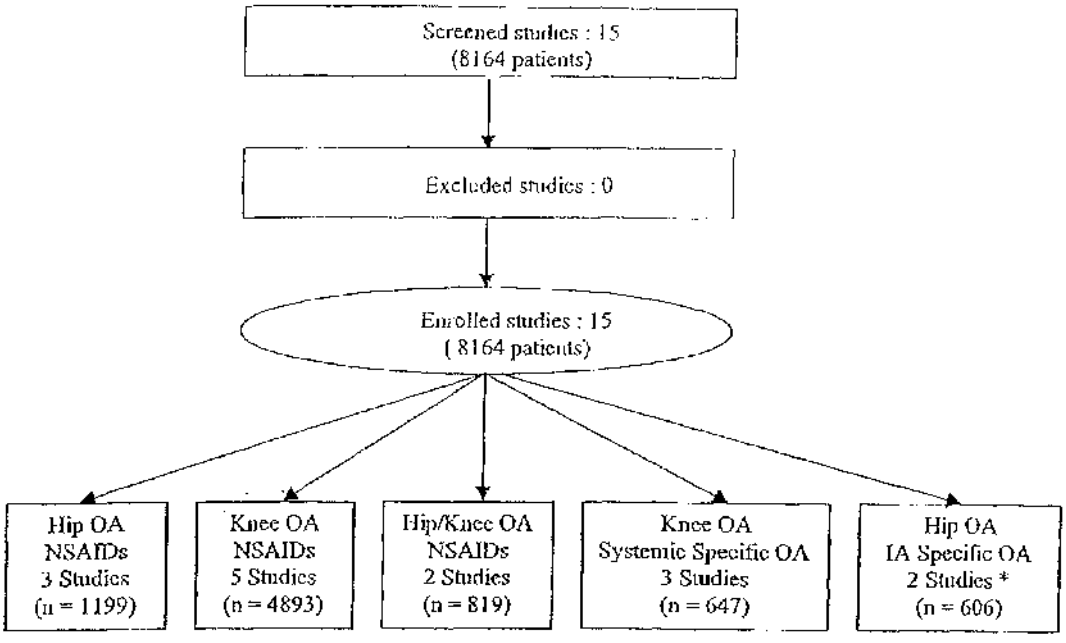


Fig. 2. Evaluated simplified sets of criteria: Scenarios C to F.



* 1 study used a non-placebo control group

Fig. 3. Flow diagram for numbers of studies and patients.

systemic specific drug, intra-articular specific drug). For each trial and each scenario, a drug company provided the number of patients and the number of responders in each treatment arm. With this information, sensitivity (percentage of patients receiving an active drug labeled as responders according to the proposed set of criteria) and specificity (percentage of patients receiving the placebo treatment labeled as non-responder according to the proposed set of criteria) could be calculated for each trial and for each drug class and joint location of interest.

The first step of the data driven approach consisted in the evaluation of the performances of the two formal scenarios (scenario A and B), using the two databases, for each category of trial studied during the elaboration step (*i.e.* hip OA–NSAIDs trials, knee OA–NSAIDs trials, knee OA–systemic specific drug trials and knee OA–intra-articular specific drug trials). In other words, we compared the following performances in the 'elaboration' and in the 'revisit' database: placebo effect (percentage of responders in the placebo group), active effect (percentage of responders in the active treatment group), treatment effect (percentage of patients improved in the active treatment group minus the percentage of patients improved in the placebo group) and the sample arm size needed to obtain the observed placebo and active treatment effects (α=0.05 and β=0.20, two tailed test).

The second step consisted in the evaluation of the above performances between the six scenarios within the revisit database. For each drug and for each OA localization, the number of patients in the active treatment group and in the placebo group, the placebo effect, the active effect, the treatment effect and the sample arm size needed to obtain the observed placebo effect and the active treatment effect were calculated. These evaluations were also calculated whatever the localization and/or whatever the treatment. Moreover, since criteria sets almost always performed optimistically well when evaluated with the same database which was used to hunt for 'optimum' scenario, we compared the performances of the scenarios C, D, E and F in the elaboration database to the performances of the scenarios A and B in the revisit database.

Experts' opinion approach: Based on the data observed and after discussion among the OMERACT VI meeting participants, a vote was conducted to select the definite OMERACT-OARSI set of criteria (one of the six evaluated scenarios).

Lastly, the sensitivity and the specificity of the selected scenario has been evaluated in the "elaboration" database (knee OA–NSAIDs trials, hip OA–NSAIDs trials). The sensitivity was defined by the percentage of NSAIDs-OA patients meeting the OMERACT-OARSI criteria. The 1-specificity was defined by the percentage of placebo-OA patients meeting the OMERACT-OARSI criteria.

Results

PATIENTS AND STUDIES

In the elaboration database, fourteen trials totaling 1886 patients were evaluated (see Ref. 5 for details). The majority of the information was on NSAIDs for knee and hip. In the revisit database, fifteen studies involving 8164 OA patients were screened (Fig. 3). None of the studies was excluded. A prospective randomized controlled study in which the control group was receiving the usual therapeutic care without true placebo (*vs.* an intra-articular OA drug) was included in the revisit database. There were no trials

Table I
Characteristics of the 15 studies included in the 'revisit' database' according to agent class

Characteristics		Drug class		
		NSAIDs	Systemic specific OA drug	Intra articular specific OA drug
Number of studies		10	3	2
Number of patients	Active drug group	5557	316	303
	Placebo or control group	1354	331	303
Study duration	(mean +/- sd; weeks)	9.3±3.8	105.3±87.7	33±9.9
Pain evaluation	WOMAC	40%	100%	50%
	VAS	60%	0	50%
	Others	0	0	0
Function evaluation	WOMAC	100%	100%	50%
	VAS	0	0	50%
	Other	0	0	0
Global assessment evaluation	Likert	40%	33%	0
	VAS	30%	66%	50%
	Other	30%		50%
Time of collection of the outcome variables	Final and baseline	100%	100%	100%
	Only final visit	0	0	0

*'Revisit' data base is the one that permitted to revisit the formal sets of responder criteria and to evaluate the simplified sets of responder criteria.
WOMAC: Western Ontario McMaster Universities Osteoarthritis index; VAS: Visual Analogic Scale.

available to examine analgesics in OA. The majority of the studies concerned NSAIDs in hip and knee OA (10 of 15).

The characteristics of study designs are summarized in Table I. The data concerned 647 patients in systemic specific OA drug trials, 606 patients in intra-articular (IA) specific OA drug trials and 6911 patients in NSAIDs trials. For NSAIDs studies, whatever the OA localization, 5557 patients received the active treatment, and 1354 the placebo. Two studies involving hip and/or knee OA without indication of the localization were included only in the "whatever the localization" calculation. To assess pain and functional disability, two tools were most often used: The visual analog scale (VAS) and the Western Ontario McMaster Universities Osteoarthritis (WOMAC) index. For global patient's assessment, the VAS and the Likert scale were mostly used.

The knee was the only OA localization of the five specific OA drug studies (systemic and intra-articular), while NSAIDs trials were conducted in both knee and hip OA.

DATA DRIVEN APPROACH RESULTS

Formal set of criteria performances: comparison between elaboration database and revisit database

Results concerning the placebo and the treatment effects are summarized in Table II. For both the propositions A and B, the variability in the placebo and the active treatment effects were quite high (from 4% to 21% in the placebo group and from 7% to 34% in the active treatment group). Based on the observed results (placebo effect and active treatment effect) in the elaboration database, the calculation of the sample size required in future NSAIDs trials in knee OA was 67 patients per arm with scenario A and 66 with scenario B.

Performances of the 6 scenarios in the revisit database according to drug class, route of administration and OA localization

The results of the evaluated performance for each scenario are summarized in Table III. The highest active treatment effect and placebo effect were observed when using scenario F, whatever the drug class and whatever the localization.

NSAIDs in Knee OA. The highest active treatment effect was observed when using scenario F (66.4%), and at variance, the lowest placebo effect was observed when using scenario B (39.1%). The treatment effect was similar whatever the scenario (19.8%, 19.3%, 19.8%, 19.5%, 19.9% and 19.8% for scenarios A, B, C, D, E and F respectively). The sample sizes "required" in future NSAID knee trials using the "revisit" data were 99 patients per arm, scenario A and 105 per arm, scenario B. Using the simplified scenarios, the sample sizes "required" were 98 per arm, scenario C, 101 per arm, scenario D, 97 per arm, scenario E and 98 per arm, scenario F.

NSAIDs in Hip OA. The highest active treatment effect was observed when using scenario F (60.8%), and at variance the lowest placebo effect was observed when using scenario A (28.9%). As observed in knee OA, the treatment effect was similar whatever the scenario (24.7%, 26.5%, 25.9%, 25.7%, 25.3% and 25.3% for scenarios A, B, C, D, E and F respectively). The sample sizes 'required' in future NSAID hip trials using the 'revisit' data were 62 patients per arm, scenario A and 55 per arm, scenario B. Using the simplified scenarios, the sample sizes 'required' were 58 per arm, scenario C, 59 per arm, scenario D and 61 per arm, scenario E and scenario F.

Systemic Specific OA drug in Knee OA. The highest active treatment effect was observed when using scenario F (49.4%), and the lowest placebo effect was observed when using scenario B (29.0%). Scenarios A and B showed the highest treatment effect (6.9% and 6.8% respectively) and the lowest sample size "required" for future systemic specific OA drug trials in knee OA (743 and 745 patients per arm respectively, *versus* 1167, 1095, 4979 and 3824 patients per arm for scenarios C, D, E and F).

Intra-articular specific OA drug in Knee OA. The highest active treatment effect was observed when using scenario F (72.9%), and the lowest placebo effect was observed when using scenario B (34.6%). The highest treatment

Table II

Performances observed with the formal sets of criteria, propositions A and B (e.g. scenarios A and B) in the 'elaboration' and in the 'revisit' databases**: placebo effect, active treatment effect, treatment effect, and variability between the two databases)*

		Formal OARSI set of criteria					
		Proposition A (pain)*			Proposition B (pain or function)*		
Trials		Elaboration**	Revisit**	(Revisit-Elaboration)	Elaboration**	Revisit**	(Revisit-Elaboration)
Knee OA Systemic Specific OA drug	Placebo effect	51%	31%	−20	50%	29%	−21
	Active treatment effect	62%	38%	−24	61%	36%	−25
	Treatment effect	11%	7%	−4	11%	7%	−4
Knee OA IA Specific OA drug	Placebo effect	47%	35%	−12	47%	35%	−12
	Active treatment effect	92%	58%	−34	91%	57%	−34
	Treatment effect	45%	23%	−22	44%	22%	−22
Hip OA NSAIDs	Placebo effect	33%	29%	−4	39%	32%	−7
	Active treatment effect	62%	54%	−8	69%	58%	−11
	Treatment effect	29%	25%	−4	30%	26%	−4
Knee OA NSAIDs	Placebo effect	27%	39%	+12	26%	39%	+13
	Active treatment effect	52%	59%	+7	51%	58%	+7
	Treatment effect	25%	20%	−5	25%	19%	−6

*See section 3 of the manuscript for detailed explanations.

**'Elaboration' database is the one that permitted to propose the formal sets of responder criteria (5); 'Revisit' database is the one that permitted to revisit the formal sets of responder criteria and to evaluate the simplified sets of responder criteria.

Table III

Performances observed with each scenario in the 'revisit' database: Percentage of patients improved in placebo and active treatment groups (i.e., placebo effect and active treatment effect), treatment effect, sample size per arm required in future trials, α=0.05, β=0.20, two-tailed, expected placebo effect=that observed with this database, expected active treatment effect=that observed with this database*

Localization Drug		Knee OA Systemic Specific OA Drug	Knee OA IA Specific OA Drug	Hip OA NSAIDs	Knee OA NSAIDs	Whatever the joint NSAIDs	Whatever the joint Whatever the systemic treatment	Whatever the joint Whatever the treatment
Scenario A	% improved in active group	38.0%**	58.4%	53.6%	59.3%	56.3%	57.2%	57.2%
	% improved in placebo group	31.1%**	35.3%	28.9%	39.5%	36.8%	35.7%	35.7%
	Treatment effect	6.9%	23.1%	24.7%	19.8%	21.5%	21.5%	21.5%
	Sample size	745	73	62	99	84	84	84
Scenario B	% improved in active group	35.8%	57.4%	56.0%	58.4%	58.7%	57.5%	57.5%
	% improved in placebo group	29.0%	34.6%	31.5%	39.1%	37.4%	35.8%	35.6%
	Treatment effect	6.8%	22.8%	26.5%	19.6%	21.3%	21.7%	21.9%
	Sample size	743	74	55	105	86	82	81
Scenario C	% improved in active group	43.7%	70.3%	60.3%	65.1%	64.7%	63.8%	63.9%
	% improved in placebo group	38.0%	42.9%	34.4%	45.3%	43.3%	42.3%	42.2%
	Treatment effect	5.7%	27.4%	25.9%	19.8%	21.4%	21.3%	21.7%
	Sample size	1167	51	58	98	84	86	82
Scenario D	% improved in active group	44.6%	70.6%	60.5%	65.4%	65.0%	63.9%	64.2%
	% improved in placebo group	38.7%	43.6%	34.8%	45.9%	43.9%	42.8%	42.9%
	Treatment effect	5.9%	27.0%	25.7%	19.5%	21.1%	21.1%	21.3%
	Sample size	1095	52	59	101	87	87	85
Scenario E	% improved in active group	47.8%	72.6%	60.5%	65.8%	65.3%	64.3%	64.7%
	% improved in placebo group	45.0%	44.5%	35.2%	45.9%	44.0%	44.2%	44.3%
	Treatment effect	2.8%	28.1%	25.3%	19.9%	21.3%	20.1%	20.4%
	Sample size	4979	48	61	97	85	96	93
Scenario F	% improved in active group	49.4%	72.9%	60.8%	66.4%	66.0%	66.1%	65.5%
	% improved in placebo group	46.2%	45.2%	35.5%	46.6%	44.5%	44.8%	44.9%
	Treatment effect	3.2%	27.7%	25.3%	19.8%	21.5%	20.3%	20.6%
	Sample size	3824	49	61	98	83	94	91

*'Revisit' database is the one that permitted to revisit the formal sets of responder criteria and to evaluate the simplified sets of responder criteria.

**Percentage of patients improved in the placebo group or in the active treatment group (*i.e.*, placebo effect and active treatment effect).

effect was observed when using scenario E (28.1%). The lowest sample size 'required' for future intra- articular specific OA drug trials in knee OA were observed when using the simplified scenarios (51, 52, 48 and 49 patients per arm for scenarios C, D, E and F respectively, *versus* 73 and 74 patients per arm for scenarios A and B).

Performances of the six scenarios in the revisit database whatever the drug class, the route of administration or the localization of OA

NSAIDs whatever the OA localization. The highest active treatment effect was observed when using scenario F (66.0%), and at variance the lowest placebo effect was observed when using scenario A (36.8%). The treatment effect was similar whatever the scenario (21.5%, 21.3%, 21.4%, 21.1%, 21.3% and 21.5% for scenarios A, B, C, D, E and F respectively). The sample size "required" for future NSAIDs trials in OA was also similar whatever the scenario (84, 86, 84, 87, 85 and 83 patients per arm for scenarios A, B, C, D, E and F respectively).

Whatever the systemic drug (i.e. systemic specific OA drugs and NSAIDs) and whatever the localization. The highest active treatment effect was observed when using scenario F (65.1%), and at variance the lowest placebo effect was observed when using scenario B (21.7%). The treatment effect was similar whatever the scenario (21.5%, 21.7%, 21.3%, 21.1%, 20.1% and 20.3% for scenarios A, B, C, D, E and F respectively). The "required" sample size was also similar whatever the scenario (84, 82, 86, 87, 90 and 94 patients per arm for scenarios A, B, C, D, F and F respectively).

Whatever the drug and whatever the localization. The highest active treatment effect was observed when using scenario F (65.5%), and at variance the lowest placebo effect was observed when using scenario B (35.6%). The treatment effect was similar whatever the scenario (21.5%, 21.9%, 21.7%, 21.3%, 20.4% and 20.6% for scenarios A, B, C, D, E and F respectively). The sample size "required" for future trials in OA was also similar whatever the scenario (84, 81, 82, 85, 93 and 91 patients per arm for scenarios A, B, C, D, E and F respectively).

EXPERTS' OPINION APPROACH RESULTS

Based on the observed results, it was considered that the data driven approach did not permit to select a specific set of criteria. However, at least two of these performances (treatment effect and required sample size) were similar whatever the scenario (A to F). These results were presented to the participants of the Osteoarthritis session of the OMERACT VI conference (Brisbane 2002). After discussion and voting, it appears that:

- The treatment effect and the required sample size were the two major characteristics to take into account in the choice of an optimal set of criteria to be used for clinical trials.
- Two other characteristics were also considered as important:

1) The definition of an improvement based not only on a relative change but also on an absolute change (scenarios A, C and E *versus* scenarios B, D and F)
2) The simplicity of the presentation: same cut-offs, set of responder criteria whatever the localization, the study drug and the route of administration (scenario A, B *versus* C, D, E and F).

Based on this preliminary discussion between experts and after a voting session, scenario D was selected (Fig. 4). It is now labeled the 'OMERACT-OARSI' set of responder criteria.

EVALUATION OF THE DIFFERENT SETS OF CRITERIA

Table IV summarizes the results of the procedure permitting the evaluation of the different scenarios. This table shows that the treatment effect was similar whatever the evaluated scenario, but for hip OA, both the sensitivity and the specificity (active treatment effect and placebo treatment effect) were higher for the scenario D.

Discussion

This study, which combined the efforts of academic researchers, representatives of the pharmaceutical industry and representatives of health agency, proposes a simplified set of responder criteria for clinical trials in OA by simplifying the initial OARSI set of criteria using a data driven and experts' opinion approach. Limitations of this study include (*i*) the absence of analgesics trials in our analysis of the improvement between active drug-treated group and placebo-treated group; (*ii*) Available trials concerned only knee or hip OA and no other OA localization; (*iii*) The collected data concerned only the core set of criteria. Drug companies provided for each trial, the percentage of responders in the active treatment group and the percentage of non-responders in the placebo group, according to each scenario. We did not have access to the individual data, neither to the percentage of responders for each domain separately (pain, function, global patient's assessment). This lack of data did not allow us to estimate if the core set of criteria was less powerful than each domain treated separately, as has been done for rheumatoid arthritis[7]; (*iv*) The cut-offs of the simplified scenarios were inspired by the formal ones. However, more specific cut-offs could not be estimated due to the lack of individual data for the 8164 OA patients.

We observed considerable variability in the results with regard to the study population (elaboration *versus* revisit database) within the formal sets of criteria. This variability could be attributed to a variability between the patients included in the two databases. However, in both of them, most of the trials have been conducted in multicenter international trials following a very similar approach concerning the inclusion and exclusion criteria (phase II and phase III trials).

In the elaboration phase of the formal OARSI set of criteria, the loss of sensitivity and specificity using identical cut-offs, whatever the localization and the study drug, did not allow to propose a simple set of criteria (similar cut-off whatever the OA localization and the study drug). The variability of the performances of these formal sets of criteria between the two databases was in contradiction with the results obtained in the elaboration phase and prompted us to further evaluate a simplification of the set of responder criteria.

The data driven conclusions are that, whatever the OA localization, the study drug or the route of administration, formal scenarios A and B had the lowest placebo effect, and scenario F had the highest active treatment effect. In

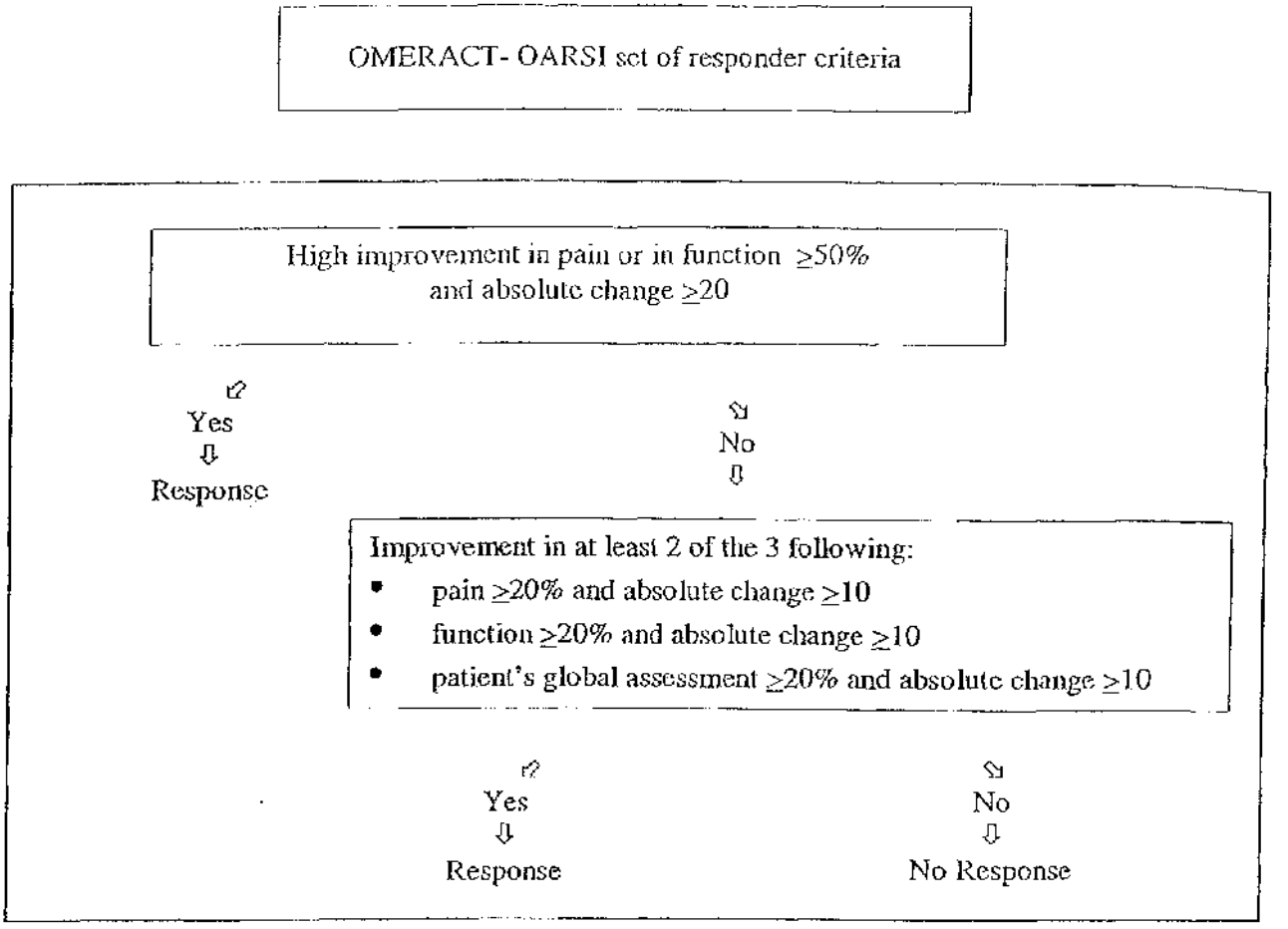


Fig. 4. OMERACT- OARSI Set of responder criteria.

Table IV
*Percentage of patients responding when the OMERACT-OARSI and formal proposition A and B criteria sets are applied to a validation data set**

Validation database	Criteria set	Knee NSAID trial		Hip NSAID trial	
		Sensitivity	Specificity	Sensitivity	Specificity
Elaboration	OMERACT-OARSI	59%	40%	72%	44%
Revisit	OARSI-Proposition A	59%	39%	54%	29%
Revisit	OARSI-Proposition B	58%	39%	58%	26%

*Sample size required per arm alpha=0.05, beta=0.20, two-tailed.Sensitivity=% responders on active drug (NSAIDs). 1-Specificity=% responders on placebo.

contrast, the treatment effect and the required sample size were quite similar whatever the scenario, ranging from 20.4% to 21.9% and from 81 to 93 patients per arm respectively.

Although the data driven approach did not allow to select any particular scenario, the simplification of the set of criteria did not result in a loss of relevant performances. Indeed, a higher active treatment effect and a higher placebo effect were observed when using the simplified scenarios in both databases. Conversely, the treatment effect and the sample size required to obtain the observed placebo and active effects were similar whatever the scenario (whether formal or simplified) in the revisit database.

According to the experts, these two performances were the most important for an optimal set of responder criteria. Although all the evaluated scenarios provided similar results for these performances, the experts' choice was scenario D (Fig. 4), which confirms the importance of:

1) A format that requires both an absolute change and a relative change.
2) A format that considers both pain and function as important domains; in certain studies, however, changes in functional disability are at least as important as changes in pain.

The observed treatment effect whatever the drug and whatever the treatment when using scenario D is 21.3%. This result is close to what is expected in OA, i.e., 20-30%[6,9].

The required sample size with scenario D whatever the drug and whatever the treatment is 85 patients per arm. This is similar to the sample size required when using the previous formal set of criteria.

In conclusion, we propose a simplified definition for symptomatic improvement in osteoarthritis. This set of criteria, approved both by the OARSI and the OMERACT committees, is at least as powerful as the previous OARSI formal set of criteria and its simplification will probably enhance its use in future OA trials.

Other studies are required on order to further validate this proposed OMERACT-OARSI set of criteria in other sets of patients suffering from osteoarthritis of different localizations and treated differently, e.g., with analgesics or non-pharmacological therapies.

Acknowledgements

The authors wish to thank the drug companies and their representatives for their active collaboration: Boehringer Ingelheim France *(Laurence Salin)*, Fidia Farmaceutici SpA *(Simonetta Piva)*, Merck *(James A. Bolognese, Elliot W. Ehrich, Ken E. Truitt)*, Negma-Lerads *(Alain Taccoen, Patrick Cohen)*, Novartis *(Ornella Della Casa Alberighi, Alan Moore)*, Pharmacia *(Carl Wallemark, Jim Lefkowitli)*, Rotta Research Laboratorium *(Lucio C. Rovati, Giampaolo Giacovelli)*.

References

1. Altman R, Brandt T, Hochberg M, Moskovitz R. Design and conduct of clinical trials of patients with osteoarthritis. Recommendations from a task force of the Osteoarthritis Research Society. Osteoarthritis Cart 1996;4:217–43.
2. Bellamy N, Kirwan J, Altman R, Boers M, Brandt K, Brooks P, *et al.* Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip and hand osteoarthritis. Results of consensus development at OMERACT III. J Rheumatol 1997;24:799–802.
3. Lequesne M, Brandt K, Bellamy N, Moskovitz R, Menkes CJ, Pelletier JP, *et al.* Guidelines for testing slow-acting drugs in osteoarthritis. J Rheumatol 1994;41(Suppl):65–71 discussion 72–3. (Published erratum appears in J Rheumatol Suppl 1994 Dec; 41: 2395.
4. Group for the respect of ethics and excellence in science (GREES): osteoarthritis section. Recommendations for the registration of drugs used in the treatment of osteoarthritis. Ann Rheum Dis 1996;55(8): 552–7.
5. Dougados M, Leclaire P, van der Heijde D, Bloch DA, Bellamy N, Altman RD. Response criteria for clinical trials on osteoarthritis of the knee and hip: a report of the Osteoarthritis Research Society International Standing Committee for Clinical Trials response criteria initiative. Osteoarthritis Cart 2000;8(6): 395–403.
6. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988;15(12):1833–40.
7. Felson DT, Anderson JJ, Boers M, Bombardier C, Furst D, Goldsmith C, *et al.* American College of Rheumatology. Preliminary definition of improvement in rheumatoid arthritis. Arthritis Rheum 1995;38(6): 727–35.
8. Altman R, Asch E, Bloch D, Bole G, Borenstein D, Brandt K, *et al.* Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Association. Arthritis Rheum 1986;29(8):1039–49.
9. Bolnot-Delmas D, Buch JP, Zeidler H, Dougados M. Ro 15-8081 in osteoarthritis of hip and knee: a double-blind placebo-controlled multicentre dose-ranging study on analgesia. Pain 1996;64(1):99–105.

EXTENDED REPORT

Evaluation of WOMAC 20, 50, 70 response criteria in patients treated with hylan G-F 20 for knee osteoarthritis

N Bellamy, M J Bell, C H Goldsmith, D Pericak, V Walker, J P Raynauld, G W Torrance, P Tugwell, R Polisson

Ann Rheum Dis 2005;64:881–885. doi: 10.1136/ard.2004.026443

See end of article for authors' affiliations

Correspondence to: Professor Nicholas Bellamy, CONROD, Faculty of Health Sciences, University of Queensland, Level 3, Mayne Medical School, Herston Road, Brisbane, Queensland 4006, Australia; nbellamy@medicine.uq.edu.au

Accepted 27 October 2004
Published Online First 25 November 2004

Objective: A secondary analysis of a previously conducted one year randomised controlled trial to evaluate the capacity of responder criteria based on the WOMAC index to detect between treatment group differences.

Methods: 255 patients with knee osteoarthritis were randomised to "appropriate care with hylan G-F 20" (AC+H) or "appropriate care without hylan G-F 20" (AC). In the original analysis, two definitions of patient response from baseline to month 12 were used: (1) at least a 20% reduction in WOMAC pain score (WOMAC 20P); (2) at least a 20% reduction in WOMAC pain score *and* at least a 20% reduction in either WOMAC function or stiffness score (WOMAC 20PFS). For this analysis, a responder was identified using 50% and 70% minimum clinically important response levels to investigate how increasing response affects the ability to detect treatment group differences.

Results: The hylan G-F 20 group had numerically more responders using all patient responder criteria. Increasing the response level from 20% to 50% detected similar differences between treatment groups (25% to 29%). Increasing the response level to 70% reduced the differences between treatment groups (11% to 12%) to a point where the differences were not significant after Bonferroni adjustment.

Conclusions: These results provide evidence for incorporating response levels (WOMAC 50) in clinical trials. While differences at the highest threshold (WOMAC 70) were not statistically detectable, an appropriately powered study may be capable of detecting differences even at this very high level of improvement.

Developments in standardisation of outcome measurement procedures for clinical trials in the treatment of osteoarthritis and rheumatoid arthritis have followed similar but not identical pathways. While measures of pain, function, and patient global assessment have been selected as core set measures for clinical trials in both of these diseases, the outcome measures in arthritis clinical trials–American College of Rheumatology (OMERACT-ACR) criteria for rheumatoid arthritis[1] differ from the OMERACT–Osteoarthritis Research Society International (OARSI) criteria for osteoarthritis,[2 3] in that the former also include measures of the number of tender and swollen joints, physician global assessment, and C reactive protein/erythrocyte sedimentation rate values. The subsequent development of responder criteria for rheumatoid arthritis[4] and osteoarthritis trials[5 6] reflects these differences in core set measures. In addition, ACR responder criteria for rheumatoid arthritis[4] are based on percentage changes on two or more variables, while OMERACT-OARSI responder criteria for osteoarthritis[6] are based on a combination of percentage and absolute changes in one or more variables. Following the development of the ACR 20 responder criteria for rheumatoid arthritis,[4] higher threshold requirements for response designation have been explored, namely ACR 50 and ACR 70 responder criteria.[7] Higher response levels have been more difficult to achieve, and between group differences in rheumatoid arthritis clinical trials have (albeit less often) been detected at these higher thresholds, requiring individual patient improvements at or above the 50% and 70% levels, respectively.[8–13]

Notwithstanding the principle of employing a combination of percentage and absolute changes of one or more variables in OMERACT-OARSI responder criteria for osteoarthritis,[6] and that of basing responder criteria on percentage change alone in OMERACT-ACR criteria for rheumatoid arthritis,[4] we undertook secondary analyses of a published randomised controlled trial[14 15] to evaluate the ability of responder criteria based on the Western Ontario and McMaster Universities (WOMAC) osteoarthritis index to detect between treatment differences. We compared the results of analyses based on WOMAC 20, WOMAC 50, and WOMAC 70 responder criteria to determine whether the application of different criteria influences data interpretation.

METHODS

Design

The analyses reported here were undertaken using the data collected in a health outcomes trial evaluating viscosupplementation with hylan G-F 20 when added to an appropriate care treatment regimen for patients with knee osteoarthritis. The detailed design of this trial and the primary analyses of the data have been published elsewhere.[14 15] Briefly, the trial was a multicentre, randomised, controlled, open label study over one year, where patients were randomised to either "appropriate care with hylan G-F 20" (AC+H) or to "appropriate care without hylan G-F 20" (AC). Appropriate care for knee osteoarthritis was defined by the guidelines for the medical management of osteoarthritis proposed by the ACR.[16] Patients in this study had symptomatic knee osteoarthritis (of mild to moderate severity) and had received previous treatment with analgesics. Appropriate care could

Abbreviations: ACR, American College of Rheumatology; WOMAC, Western Ontario and McMaster Universities osteoarthritis index

include treatment with analgesics, non-steroidal anti-inflammatory drugs (NSAIDs), corticosteroid injections, supportive measures such as education and counselling, weight loss, joint rest, application of heat or ice, use of devices, physical therapy, arthroscopy, and total joint replacement. Patients randomised to the AC+H group could receive more than one course of hylan G-F 20 treatment in the study knee (the knee that was most symptomatic or with the predominant musculoskeletal problem) if medically warranted, and could receive bilateral treatment if their contralateral knee was affected. Retreatment was provided when persistent pain recurred, with a minimum of four weeks between courses of hylan G-F 20. Patients were assessed by the clinical investigator at baseline and at 12 months. Follow up assessments were completed by telephone at months 1, 2, 4, 6, 8, 10, and 12. The study protocol and informed consent form were approved by the relevant ethics committees for the sites. Informed consent was obtained from each patient.

Outcome measures

The WOMAC Likert 3.0 is a self administered, disease specific health related quality of life instrument that asks the patients questions concerning the study knee. It produces one aggregate total score and scores for three subscales: pain, stiffness, and physical functioning. A higher score for each subscale corresponds to a worse condition. The pain subscale includes five questions on the degree of pain experienced with certain positions and activities (for example, sitting or lying), with the subscore varying from 0 to 20. The function subscale includes 17 questions on the degree of difficulty experienced while completing activities (for example, descending stairs); the subscore varies from 0 to 68. The stiffness subscale includes two questions on severity of stiffness (that is, after first awakening, and later in the day), with the subscore varying from 0 to 8. For every question in the WOMAC, patients rate their pain, stiffness, or function using five ordinal responses: none, mild, moderate, severe, and extreme. The WOMAC was completed in the office at baseline and by telephone at months 1, 2, 4, 6, 8, 10, and 12.

Table 1 Demographic variables and disease status at baseline

Baseline measure	AC+H (n = 127)	AC (n = 128)
Age (years)	62.6 (9.4)	63.5 (10.5)
Body mass index (kg/m^2)	32.1 (6.0)	32.9 (7.2)
Duration (years) of OA symptoms		
Study knee	9.0 (9.5)	9.9 (9.7)
Other knee	7.4 (8.8)	8.3 (9.3)
Sex: female	86 (68%)	93 (73%)
Previous treatment for OA of the knee(s)		
Acetaminophen	100 (79%)	109 (85%)
NSAIDs	120 (94%)	110 (86%)
Previous surgery, study knee	40 (31%)	39 (30%)
Previous surgery, other knee	40 (31%)	23 (18%)
Radiology grading within 1 year (central grading*)		
Not reported	0 (0%)	1 (1%)
Grade 0	4 (3%)	4 (3%)
Grade I	17 (13%)	11 (9%)
Grade II	32 (25%)	33 (26%)
Grade III	49 (39%)	37 (29%)
Grade IV	25 (20%)	42 (33%)
OA at baseline		
Other knee affected	109 (86%)	108 (84%)
Other joints affected	95 (75%)	87 (68%)

Values are mean (SD) or n (%).
*Radiology grading is based on central grading, which may have differed from the site investigator's determination for patient eligibility.
AC, appropriate care; AC+H, appropriate care + hylan G-F 20; NSAID, non-steroidal anti-inflammatory drug; OA, osteoarthritis.

In the original study analysis,[14] the primary effectiveness measure was the mean change in the WOMAC pain subscore in the study knee from baseline to month 12. Secondary effectiveness measures included two definitions of a responder that incorporated a minimum clinically important response level of at least 20%. These measures were defined as the percentage of patients improved by month 12 (compared with baseline) using different combinations of the WOMAC subscales as follows: (1) at least a 20% improvement from baseline in the WOMAC pain score in the study knee (WOMAC 20P); (2) at least a 20% improvement from baseline in the WOMAC pain score in the study knee *and* at least a 20% improvement from baseline in either the function score or the stiffness score (WOMAC 20PFS).

Alternative patient responder criteria

Alternative patient responder criteria were examined in this analysis. Recent trials in rheumatoid arthritis have used higher threshold levels to define a patient responder, to "raise the bar" and define rheumatoid arthritis improvements by more substantial changes in core set measures.[7] While the 20% minimum clinically important response level used to define a patient responder in our original study was able to discriminate between the AC+H and AC treatment groups, we increased the minimum clinically important response levels to 50% and 70%.

These new criteria incorporate the pain, function, and stiffness subscores from the WOMAC, identical to the original secondary effectiveness measures. For the 50% minimum clinically important response level, the definitions were: (1) at least a 50% improvement from baseline in the WOMAC pain score in the study knee (WOMAC 50P); and (2) at least a 50% improvement from baseline in the WOMAC pain score in the study knee *and* at least a 50% improvement from baseline in either the function score or the stiffness score (WOMAC 50PFS). Similarly, for the 70% minimum clinically important response level, the definitions were: (1) at least a 70% improvement from baseline in the WOMAC pain score in the study knee (WOMAC 70P); and (2) at least a 70% improvement from baseline in the WOMAC pain score in the study knee *and* at least a 70% improvement from baseline in either the function score or the stiffness score (WOMAC 70PFS). These responder criteria can be collectively termed the WOMAC 20, WOMAC 50, and WOMAC 70 criteria.

Differences between treatment groups

Discriminant validity, which has been defined as the ability of a measure to distinguish clinically important differences between treatment groups,[17] was evaluated using these responder criteria. We hypothesised that when increasing the threshold for defining patient improvement, the number of patients classified as responders in both treatment groups would decrease. However, it is unclear how this would affect

Table 2 WOMAC scores at baseline

Baseline measure	AC+H (n = 127)	AC (n = 127)*
WOMAC pain (0 to 20)†	11.4 (2.7)	11.9 (2.9)
WOMAC stiffness (0 to 8)†	5.1 (1.5)	5.1 (1.4)
WOMAC function (0 to 68)†	39.5 (9.3)	40.2 (9.3)

Values are mean (SD).
*One patient in the AC group did not have a baseline WOMAC questionnaire completed and was thus not included in the analysis.
†The higher the score, the worse the problem.
AC, appropriate care; AC+H, appropriate care + hylan G-F 20; WOMAC, Western Ontario and McMaster Universities osteoarthritis index.

the overall treatment group differences for each patient improved definition.

In the original study, a 20% difference between treatment groups for the primary and secondary effectiveness measures was established a priori by the steering committee as the minimum clinically important difference based in part on previous research.[18] In addition, a 20% improvement was the minimum clinically important improvement from baseline to month 12 for each patient who was classified as a responder.

Statistical methods

Data from the locked study database were analysed using SAS version 8.2. Multivariable logistic analyses were undertaken for each of the responder criteria that incorporated different minimum clinically important response levels. Patient were classified responders if they improved according to the criteria outlined in the definition from baseline to month 12. The hypothesis tested was whether AC+H was superior to AC when the responder criteria were applied.

All analyses were adjusted for design variables—that is, baseline value of the variable being analysed, site, blocking by site, body mass index, and baseline WOMAC total score. The type 1 experiment-wise error rate was controlled for by distributing α over all six response levels (that is, WOMAC 20P and 20PFS; WOMAC 50P and 50PFS; WOMAC 70P and 70PFS) using the Bonferroni adjustment of $\alpha/6$ (α for each comparison = 0.05/6 = 0.0083). The original secondary effectiveness measures are provided for comparison with the patient improved definitions which incorporate higher minimum clinically important response levels.

All patients were included in the intent to treat group for all analyses as described earlier.[14] The hot deck method was used to impute data for all effectiveness measures as described earlier.[14]

RESULTS

Patients

In the trial, 128 patients were randomised to receive appropriate care and 127 patients to receive appropriate care with hylan G-F 20. In all, 24 patients dropped out of the study (21 in the AC group, three in the AC+H group). One patient in the AC group did not have a baseline WOMAC questionnaire completed and thus was not included in the analysis. Descriptive statistics comparing demographic variables, baseline disease characteristics, and baseline outcome measures (that is, WOMAC pain, function, and stiffness subscores) are given in tables 1 and 2. Overall, treatment groups were similar for demographics, disease characteristics, and osteoarthritis treatments used at baseline. However, 20% of patients in the AC+H group and 33% in the AC group had grade IV osteoarthritis, as subsequently determined by central radiological grading. WOMAC scores for pain, stiffness, and function were similar between groups.

Knee osteoarthritis treatment

All patients except one in the AC+H group had at least one course of hylan G-F 20 in their study knee, and 53 (42%) had at least one course in their contralateral knee. Forty five patients (38%) in the AC+H group received a second course of hylan G-F 20 in their study knee, and three received a third course in their study knee. Twenty patients (16%) in the AC+H group received a second course in their contralateral knee. More patients in the AC group than in the AC+H group reported corticosteroid injections in the study knee (70% *v* 14%) or in the contralateral knee (27% *v* 6%) (both $p<0.0001$). There were also more patients in the AC group taking NSAIDs for any knee (79% *v* 65%) ($p = 0.0062$), and other drugs (20% *v* 10%) (for example, opioid analgesics, anti-inflammatory agents) for any knee ($p = 0.0216$). There were no significant differences between the groups in the use of concomitant drug treatment for overall osteoarthritis. Further details of the knee osteoarthritis treatment can be found in the original study results.[14]

Effectiveness

The results for the original secondary effectiveness measures and new responder criteria are given in table 3. They showed that for both the original secondary effectiveness measure and the alternative patient responder criteria, the percentage of responders was greater in the AC+H group than in the AC group. The treatment group differences were significant at the 0.0083 level ($\alpha/6 = 0.05/6$) for the 20% and 50% minimum clinically important response levels (adjusted using Bonferroni correction) and exceeded the required 20% difference established a priori as the minimum clinically important difference between treatment groups (25% to 29%). When the minimum clinically important response level increased to 70%, the treatment group differences were approximately one half the size (that is, 11% to 12%) of the differences found with the 20% and 50% levels, and did not reach statistical significance after Bonferroni correction. Within each minimum clinically important response level, the treatment group differences were similar regardless of whether the WOMAC pain scores, or all of the WOMAC pain, function, and stiffness scores, were incorporated into the

Table 3 Number (%) of patient responders using WOMAC 20, 50, and 70 minimum clinically important response levels

Patient responder definitions	Treatment group: AC+H (n = 127)	Treatment group: AC* (n = 127)	Difference [(AC+H) − AC)]	p Value
Original effectiveness measures				
WOMAC 20P†	87 (69%)	51 (40%)	29%	0.0001‡
WOMAC 20PFS†	79 (62%)	45 (35%)	27%	0.0001‡
Alternative responder criteria				
WOMAC 50P	53 (42%)	20 (16%)	26%	<0.0001‡
WOMAC 50PFS	43 (34%)	12 (9%)	25%	<0.0001‡
WOMAC 70P	26 (20%)	10 (8%)	12%	0.0118
WOMAC 70PFS	20 (16%)	6 (5%)	11%	0.0100

*One patient in the AC group did not have a baseline WOMAC questionnaire completed and thus was not included in the analysis.
†Secondary effectiveness measures from main study.
‡Analysis controlled for the type I experiment-wise error rate by distributing α over all response levels using Bonferroni adjustment $\alpha/6 = 0.05/6 = 0.0083$.
AC, appropriate care; AC+H = appropriate care + hylan G-F 20.

definition (for example, 29% for WOMAC 20P, 27% for WOMAC 20PFS).

The percentage of patients classified as responders decreased for both treatment groups as response levels increased from 20% to 70%, and with the more stringent definition incorporating pain, function, and stiffness within each response level. Considering the AC+H group, when moving from the lower to the higher response level for pain only (that is, WOMAC 20P to WOMAC 70P), the percentage of responders decreased from 69% to 20%. Similarly, when increasing the response levels with the more stringent criteria incorporating pain and either function or stiffness (that is, WOMAC 20PFS to WOMAC 70PFS), a similar decrease was observed in the AC+H group (62% to 16%). For the AC group, large decreases were also found when response levels increased for the criteria incorporating only pain (40% to 8%), and the more stringent criteria incorporating pain and either function or stiffness (35% to 5%).

When comparing the AC+H group and the AC group for all responder criteria, the results show that the percentage of responders in the AC+H group relative to the AC group was generally greater for criteria that incorporate the higher minimum clinically important response levels. For example, for the WOMAC 70PFS criterion, the percentage of responders in the AC+H group was approximately three times the percentage of responders in the AC group (that is, 16% *v* 5%). This is in comparison to the WOMAC 20PFS criterion where the percentage of responders in the AC+H group was less than twice the percentage of responders in the AC group (that is, 62% *v* 35%). This pattern was also observed with the criteria incorporating pain (WOMAC 20P to WOMAC 70P).

DISCUSSION

Traditional methods of carrying out between group comparisons of clinical trials data are often based on the analysis of continuous variables. These provide an appreciation of the magnitude and variation of group effects but do not usually translate into an understanding of the degree of improvement experienced by individual patients. In contrast, responder criteria, while being reductionist from a group standpoint, are capable of categorising individual patients according to whether they achieve levels of improvement at or above prespecified response thresholds. Response thresholds have generally been established a priori either to reflect a clinically important difference at an individual level, or on the basis of differentiating most efficiently between an active treatment and a placebo control.[4–6] In the case of the effectiveness measures used in the original study,[14] these were proposed during protocol development at a time when there was no precedent to follow, but 20% was considered by the development group to represent a minimum clinically important difference, and one that was of the same order of magnitude as the previously published ACR 20 criteria[4] for rheumatoid arthritis. The OARSI responder criteria[5] were developed during the execution of the protocol, and the OMERACT-OARSI responder criteria[6] were developed following completion of the study, but neither were available at study initiation. It is of interest therefore that WOMAC 20 and WOMAC 50 responder criteria, based on pain only or on the pain, stiffness, and function subscales, yield statistically detectable between group differences of the order of 25% to 29%, with percentage response at WOMAC 20 being slightly higher numerically than at WOMAC 50. Indeed this approach to the analysis provides additional confirmation of the clinical and statistical superiority of adding hylan G-F 20 to appropriate care regimens in the treatment of knee osteoarthritis. While differences at the highest threshold level (WOMAC 70) were not statistically detectable after Bonferroni correction and may be more difficult to attain, an appropriately powered study could be capable of detecting differences in patient attainment rates at even this very high level of percentage improvement. This approach to dissecting the differential therapeutic response can be considered complementary to other responder criteria and should not be considered as replacing more traditional methods. Whether these observations can be generalised to patients with either more or less severe symptoms requires further study.

A potential limitation of response criteria based on percentage change is that the accompanying absolute change can differ markedly. Thus a 20% improvement for a patient with a baseline score of 20 normalised units (NU) (0–100 NU scale) is 4 units, whereas a 20% improvement for a patient with a baseline score of 75 NU is 15 units. Furthermore, in a comparison of outcome measures in rheumatoid arthritis clinical trials, Anderson *et al*[19] noted that measures based on continuous data provided better responsiveness than the ACR 20 or disease activity score. Nevertheless, response criteria based on percentage change offer simplicity, some comparability with OMERACT-ACR criteria for rheumatoid arthritis, and an opportunity to review the number of patients who attain or exceed a prespecified threshold. While the use of responder criteria may have a negative impact on statistical power for clinical trials applications, it does provide a novel approach to outcome measurement which may facilitate the use of quantitative measurement procedures in clinical practice applications.

The results of this analysis provide evidence for the capacity of WOMAC 20, 50, and 70 responder criteria to detect clinically important and statistically significant differences between two active treatment groups in a pragmatic randomised trial. In particular we have observed—as judged by each of the four criteria sets with Bonferroni correction and by all six criteria sets without correction—that significantly more patients in the AC+H group achieved responder status than in the AC group. This approach, based on percentage improvement in pain alone or in pain and either stiffness or function, allows reviewers and consumers to discern how many patients experienced a clinically important reduction in symptom severity. Given that the analytical strategy is individualised, this approach may have important implications for monitoring patients in routine clinical care and facilitating evidence based therapeutic decision making and shared goal setting in various health care environments.

ACKNOWLEDGEMENTS

We thank Genzyme Corporation for funding the study. The study agreement with Genzyme Corporation gave the investigators independence to publish regardless of the results. We would also like to thank the clinical investigators who enrolled patients and are listed in the main manuscript.

Authors' affiliations

N Bellamy, CONROD, Faculty of Health Sciences, The University of Queensland, Brisbane, Queensland, Australia
M J Bell, Rheumatology and Clinical Epidemiology, Sunnybrook & Women's Health Sciences Centre, University of Toronto, Toronto, Ontario, Canada
C H Goldsmith, G W Torrance, Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada
D Pericak, V Walker, Innovus Research Inc, Burlington, Ontario, Canada
J P Raynauld, Department of Medicine, Montreal Institute of Rheumatology, Montreal, Quebec, Canada
P Tugwell, Centre for Global Health, Institute of Population Health, University of Ottawa, Ottawa, Ontario, Canada
R Polisson, Genzyme Corporation, Cambridge, Massachusetts, USA

REFERENCES

1 **Felson DT**, Anderson JJ, Boers M, Bombardier C, Chernoff M, Fried B, *et al*. The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. The Committee on Outcome Measures in Rheumatoid Arthritis Clinical Trials. *Arthritis Rheum* 1993;36:729–40.

2 **Osteoarthritis Research Society (OARS) Task Force**. Design and conduct of clinical trials of patients with osteoarthritis: recommendations from a task force of the Osteoarthritis Research Society. *Osteoarthritis Cart* 1996;4:217–43.

3 **Bellamy N**, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P, *et al*. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip and hand osteoarthritis. Consensus development in OMERACT III. *J Rheumatol* 1997;24:799–802.

4 **Felson DT**, Anderson JJ, Boers M, Bombardier C, Furst D, Goldsmith CH, *et al*. American College of Rheumatology. Preliminary definition of improvement in rheumatoid arthritis. *Arthritis Rheum* 1995;38:727–35.

5 **Dougados M**, LeClaire P, van der Heijde D, Bloch DA, Bellamy N, Altman RD, *et al*. Response criteria for clinical trials on osteoarthritis of the knee and hip: a report of the Osteoarthritis Research Society International Standing Committee for Clinical Trials Response Criteria Initiative. *Osteoarthritis Cart* 2000;8:395–403.

6 **Pham T**, van der Heijde D, Lassere M, Altman RD, Anderson JJ, Bellamy N, *et al*. Outcome variables for osteoarthritis clinical trials: the OMERACT-OARSI set of responder criteria. *J Rheumatol* 2003;30:1648–54.

7 **Felson DT**, Anderson JJ, Lange ML, Wells G, LaValley MP. Should improvement in rheumatoid arthritis clinical trials be defined as fifty percent or seventy percent improvement in core set measures, rather than twenty percent? *Arthritis Rheum* 1998;41:1564–70.

8 **Kalden JR**, Schattenkirchner M, Sorensen H, Emery P, Deighton C, Rozman B, *et al*. The efficacy and safety of leflunomide in patients with active rheumatoid arthritis: a five-year follow up study. *Arthritis Rheum* 2003;48:1513–20.

9 **Hochberg MC**, Tracy JK, Hawkins-Holt M, Flores RH. Comparison of the efficacy of the tumour necrosis factor alpha blocking agents adalimumab in rheumatoid arthritis. *Ann Rheum Dis* 2003;62(suppl 2):13–16.

10 **Kremer JM**, Weinblatt ME, Bankhurst AD, Bulpitt KJ, Fleischmann RM, Jackson CG, *et al*. Etanercept added to background methotrexate therapy in patients with rheumatoid arthritis: continued observations. *Arthritis Rheum* 2003;48:1493–9.

11 **Cohen S**, Hurd E, Cush J, Schiff M, Weinblatt ME, Moreland LW. Treatment of rheumatoid arthritis with anakinra, a recombinant human interleukin-1 receptor antagonist, in combination with methotrexate: results of a twenty-four week, multicenter, randomized, double-blind, placebo-controlled trial. *Arthritis Rheum* 2002;46:614–24.

12 **Weisman MH**, Moreland LW, Furst DE, Weinblatt ME, Keystone EC, Paulus HE, *et al*. Efficacy, pharmacokinetic, and safety assessment of adalimumab, a fully human anti-tumor necrosis factor-alpha monoclonal antibody, in adults with rheumatoid arthritis receiving concomitant methotrexate: a pilot study. *Clin Ther* 2003;25:1700–21.

13 **Mease PJ**, Goffe BS, Metz J, VanderStoep A, Finck B, Burge DJ. Etanercept in the treatment of psoriatic arthritis and psoriasis: a randomized trial. *Lancet* 2000;356:385–90.

14 **Raynauld JP**, Torrance GW, Band PA, Goldsmith CH, Tugwell P, Walker V, *et al*. A prospective, randomized, pragmatic health outcomes trial evaluating the incorporation of hylan G-F 20 into the treatment paradigm for patients with knee OA: clinical results. *Osteoarthritis Cart* 2002;10:506–17.

15 **Torrance G**, Raynaud JP, Walker V, Goldsmith CH, Bellamy N, Band P, *et al*. A prospective randomized health outcomes trial evaluating the incorporation of hylan GF-20 into the treatment paradigm for patients with knee osteoarthritis: economic results. *Osteoarthritis Cart* 2002;10:518–27.

16 **Hochberg MC**, Altman RD, Brandt KD, Clark BM, Dieppe PA, Griffin MR, *et al*. Guidelines for the medical management of osteoarthritis. Part II. Osteoarthritis of the knee. *Arthritis Rheum* 1995;38:1541–6.

17 **Bombardier C**, Tugwell P. A methodological framework to develop and select indices for clinical tasks: statistical and judgmental approaches. *J Rheumatol* 1982;9:753–7.

18 **Goldsmith CH**, Boers M, Bombardier C, Tugwell P. Criteria for clinically important changes in outcomes: development, scoring, and evaluation of rheumatoid arthritis patient and trial profiles. OMERACT Committee. *J Rheumatol* 1993;20:561–5.

19 **Anderson JJ**, Bolognese JA, Felson DT. Comparison of rheumatoid arthritis clinical trial outcome measures. *Arthritis Rheum* 2003;48:3031–8.

OsteoArthritis and Cartilage (2005) **13**, 104–110

doi:10.1016/j.joca.2004.10.016

Osteoarthritis and Cartilage

International Cartilage Repair Society

OARSI OSTEOARTHRITIS

The effectiveness of hylan G-F 20 in patients with knee osteoarthritis: an application of two sets of response criteria developed by the OARSI and one set developed by OMERACT–OARSI[1]

Nicholas Bellamy M.D.†*, Mary J. Bell M.D.‡, Charlie H. Goldsmith Ph.D.§, Dan Pericak M.Math.‖, Valery Walker M.Sc.‖, Jean-Pierre Raynauld M.D.¶, George W. Torrance Ph.D.§‖, Peter Tugwell M.D.†† and Richard Polisson M.D.‡‡

† *CONROD, Faculty of Health Sciences, The University of Queensland, Brisbane, Queensland, Australia*
‡ *Rheumatology and Clinical Epidemiology, Sunnybrook & Women's Health Science Centre, University of Toronto, Toronto, ON, Canada*
§ *Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, ON, Canada*
‖ *Innovus Research Inc., Burlington, ON, Canada*
¶ *Montreal Institute of Rheumatology, Montreal, PQ, Canada*
†† *Centre for Global Health, Institute of Population Health, University of Ottawa, Ottawa, ON, Canada*
‡‡ *Genzyme Corporation, Cambridge, MA, USA*

Summary

Objective: Secondary analyses of a previously conducted 1-year randomized controlled trial were performed to assess the application of responder criteria in patients with knee osteoarthritis (OA) using different sets of responder criteria developed by the Osteoarthritis Research Society International (OARSI) (Propositions A and B) for intra-articular drugs and Outcome Measures in Arthritis Clinical Trials (OMERACT)–OARSI (Proposition D).

Methods: Two hundred fifty-five patients with knee OA were randomized to "appropriate care with hylan G-F 20" (AC+H) or "appropriate care without hylan G-F 20" (AC). A patient was defined as a responder at month 12 based on change in Western Ontario and McMaster Universities Osteoarthritis Index pain and function (0–100 normalized scale) and patient global assessment of OA in the study knee (at least one-category improvement in very poor, poor, fair, good and very good). All propositions incorporate both minimum relative and absolute changes.

Results: Results demonstrated that statistically significant differences in responders between treatment groups, in favor of hylan G-F 20, were detected for Proposition A (AC+H = 53.5%, AC = 25.2%), Proposition B (AC+H = 56.7%, AC = 32.3%) and Proposition D (AC+H = 66.9%, AC = 42.5%). The highest effectiveness in both treatment groups was observed with Proposition D, whereas Proposition A resulted in the lowest effectiveness in both treatment groups. The treatment group differences always exceeded the required 20% minimum clinically important difference between groups established *a priori*, and were 28.3%, 24.4% and 24.4% for Propositions A, B and D, respectively.

Conclusion: This analysis provides evidence for the capacity of OARSI and OMERACT–OARSI responder criteria to detect clinically important statistically detectable differences between treatment groups.

Key words: Effectiveness, Hylan G-F 20, Osteoarthritis, Randomized controlled trial, Responder criteria.

Abbreviations: AC, Appropriate care without hylan G-F 20; AC+H, Appropriate care with hylan G-F 20; ACR, American College of Rheumatology; NSAIDs, Nonsteroidal anti-inflammatory drugs; NU, Normalized units; OA, Osteoarthritis; OARSI, Osteoarthritis Research Society International; OMERACT, Outcome Measures in Arthritis Clinical Trials; SD, Standard deviation; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index

Introduction

Several different approaches to developing a definition for a "responder" have been proposed for osteoarthritis (OA) clinical studies[1–5]. A conceptual approach to generally categorizing various types of clinical difference has been advanced by Beaton and colleagues[6]. A systematic approach, based on the statistical analysis of data from several previously completed placebo-controlled clinical trials and multistakeholder consensus, has been conducted by a working group of the Osteoarthritis Research Society International (OARSI), and resulted in the formulation of OARSI responder criteria A and B[7]. The original criteria were different for different joints and classes of interventions, Proposition A placing priority on high levels of pain relief, while Proposition B permitted high relief in pain or

[1]Source(s) of support: This study was funded by Genzyme Corporation.
*Address correspondence and reprint requests to: Prof. Nicholas Bellamy, CONROD, Faculty of Health Sciences, The University of Queensland, Level 3, Mayne Medical School, Herston Road, Brisbane, Queensland 4006, Australia. Tel: 61-7-3365-5558; Fax: 61-7-3346-4603; E-mail: nbellamy@medicine.uq.edu.au
Received 27 May 2004; revision accepted 22 October 2004.

better function. Subsequently, the Outcome Measures in Arthritis Clinical Trials (OMERACT)–OARSI collaborative exercise evaluated six scenarios and developed agreement around Proposition D as a simplified set of responder criteria applicable to different joints and types of interventions[8]. In each of the aforementioned criteria sets, patients are categorized as responders based on a combination of absolute and percentage change on one or more OARSI core set clinical measures of pain, physical function and patient global assessment[9,10]. To assess the application of responder criteria in OA patients, we have performed secondary analyses of a previously published clinical study[1,11] to evaluate the capacity of responder criteria to detect between-treatment differences in a randomized clinical trial. In particular, we have compared the results of analyses based on OARSI Proposition A, OARSI Proposition B and OMERACT–OARSI Proposition D to ascertain whether the application of different criteria influences data interpretation.

Methods

DESIGN

The analyses reported here were performed using the data collected in a health outcomes trial evaluating viscosupplementation with hylan G-F 20 when added to an appropriate care treatment paradigm for patients with knee OA. The detailed design of this trial and the primary analyses of the data have been published elsewhere[1]. Briefly, the trial was a multicenter, randomized, controlled, open-label study of 1-year duration, where patients were randomized to either "appropriate care with hylan G-F 20" (AC+H) or "appropriate care without hylan G-F 20" (AC). Appropriate care for knee OA was defined by the Guidelines for the Medical Management of Osteoarthritis of the Knee proposed by the American College of Rheumatology (ACR)[12]. Patients in this study had symptomatic knee OA (mild to moderate severity) and had received prior treatment with analgesics. Appropriate care could include medications such as analgesics, nonsteroidal anti-inflammatory drugs (NSAIDs), corticosteroid injections, supportive measures such as education and counseling, weight loss, joint rest, application of heat or ice, and use of devices, physical therapy, arthroscopy, and total joint replacement. Patients randomized to the AC+H group could receive more than one course of hylan G-F 20 treatment in the study knee (knee most symptomatic or with the most predominant musculoskeletal problem) if medically warranted, and could receive bilateral treatment if their contralateral knee was affected. Re-treatment was provided when persistent pain recurred, with a minimum of 4 weeks between courses of hylan G-F 20. The protocol did not allow hylan G-F 20 treatment for OA in joints other than the knee. Patients were assessed by the clinical investigator at baseline and 12 months. Follow-up assessments were completed by telephone at months 1, 2, 4, 6, 8, 10 and 12.

OUTCOME MEASURES

The Western Ontario and McMaster Universities (WOMAC) Osteoarthritis Index Likert 3.0 is a self-administered, disease-specific health-related quality of life instrument that asks the patients questions concerning the study knee. It produces one aggregate total score (minimum: 0; maximum: 30) and scores for three subscales: pain, stiffness, and physical functioning. A greater score for each subscale corresponds to a worsening condition. The pain subscale includes five questions regarding the degree of pain experienced with certain positions and activities (e.g., sitting or lying), with a pain sub-score varying from 0 to 20. The function subscale includes 17 questions regarding the degree of difficulty experienced while completing activities (e.g., descending stairs). The function sub-score varies from 0 to 68. The stiffness subscale includes two questions on severity of stiffness (i.e., after first awakening, and later in the day) with a sub-score varying from 0 to 8. For every question in the WOMAC, patients rate their pain, stiffness, and function using five ordinal responses: none, mild, moderate, severe, and extreme.

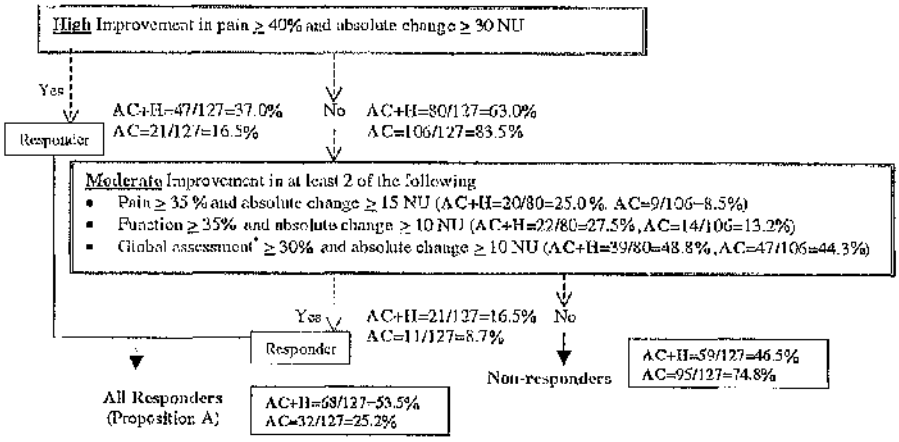
Patient global assessments were also measured. The questions that were completed by the patient were developed specifically for the original study, as no standardized questions were available[13]. Patients were asked: "how has the osteoarthritis in your study knee been during the past 4 weeks?" This question was also asked for OA in all affected joints, and for overall health. Patients answered these questions using five ordinal responses: very good, good, fair, poor and very poor. The WOMAC and the patient global assessments were completed in the office at baseline and by telephone at follow-up assessments.

In the original study analysis[1], the primary measure of effectiveness was the mean change in the WOMAC pain sub-score in the study knee from baseline. Secondary effectiveness measures were the percentage of patients improved at termination (compared to baseline) using different combinations of the WOMAC subscales to define an improved patient as follows: (1) at least 20% improvement from baseline in the WOMAC pain score in the study knee; and (2) at least 20% improvement from baseline in the WOMAC pain score in the study knee and at least 20% improvement from baseline in either function score or stiffness score. The next section describes alternative response criteria that were developed to provide a more thorough characterization of who in general improves with OA[7,8].

OARSI AND OMERACT RESPONSE CRITERIA

Alternative effectiveness measures were examined in this study. These measures were the percentage of patients improved at termination (month 12) compared to baseline using the three new responder criteria. Two of these criteria were developed by the OARSI (Propositions A and B) specifically to the treatment of knee OA with intra-articular drugs[7]. The third was developed by the OMERACT–OARSI[8] task force and is a simplified set of criteria (Proposition D) that is independent of drug and localization of OA (see Figs. 1–3). According to all propositions, a patient is defined as a responder, if either a 'high' response in pain (Proposition A) or a high response in pain or function (Propositions B and D) or a 'moderate' response in at least two domains (of pain, function, global assessment) is achieved.

The following measures from our study were used to generate inputs for determining fulfillment of the criteria, as outlined in Figs. 1–3. Pain and function were measured by the WOMAC pain and function subscales, at month 12 and compared to baseline. The pain and function subscales of the WOMAC were converted to a 0–100 normalized units (NU) scale to correspond with the specific criteria outlined in Propositions A, B and D. Patient global assessment of OA in the study knee was measured using five ordinal responses: very good, good, fair, poor, and very poor.



*Moderate improvement: In global assessment in these results was considered as at least a one-category, shift (improvement) on a 5 point ordinal response scale (i.e. very poor, poor, fair, good and very good);

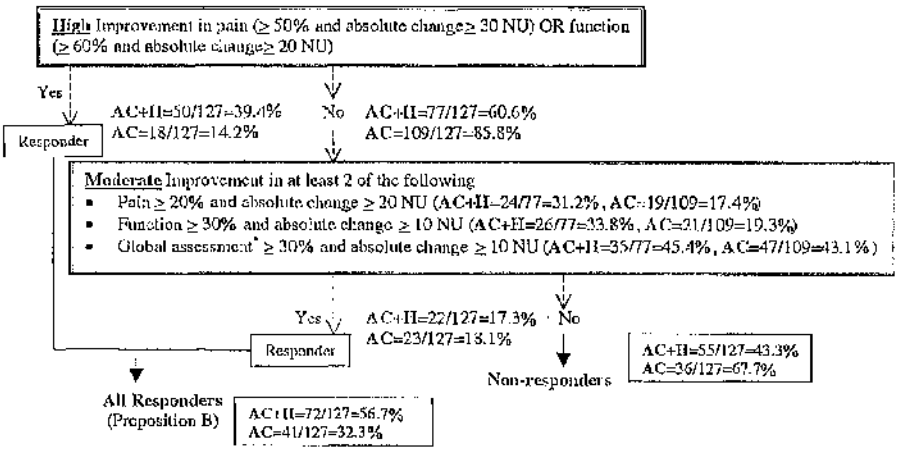
**Proposition A criteria for knee intra-articular specific drug;

NU = Normalized Units;
AC+H = Appropriate Care + Hylan G-F 20;
AC = Appropriate Care.

Fig. 1. Percentage of patients fulfilling specific criteria and overall response criteria for Proposition A**.

The problem with such a scale is that calculating a percentage change (e.g., at least 30%) or absolute change (e.g., at least 10 NU on a 0—100 scale) in the spirit of the propositions is not meaningful since the global scale is discrete with only five categories. The proposition criteria are suited more towards continuous scales, or scales that are approximately continuous (e.g., WOMAC). Normalizing the global assessment scale to 100 NU resulted in a one-category improvement shift being equivalent to 25 NU, thus never allowing the proposition criteria to be exactly fulfilled (i.e., absolute change of 10 NU required for all propositions). Thus, given the ordinal nature of the patient global assessment scale, 'moderate improvement' was not defined according to the proposition criteria; rather it was operationally defined as at least a one-category improvement shift in response (e.g., good to very good).

Based on previous research[14], a 20% difference between treatment groups for the original measures of effectiveness was established *a priori* by the steering committee and investigators as the minimum clinically important difference. This difference was applied to the new OARSI responder criteria effectiveness measures.



*Moderate improvement in global assessment in these results was considered as at least a one-directional category shift (improvement) on a 5 point ordinal response scale (i.e. very poor, poor, fair, good and very good).

**Proposition B criteria for knee intra-articular specific drug.

NU = Normalized Units;
AC+H = Appropriate Care + Hylan G-F 20;
AC = Appropriate Care.

Fig. 2. Percentage of patients fulfilling specific criteria and overall response criteria for Proposition B**.

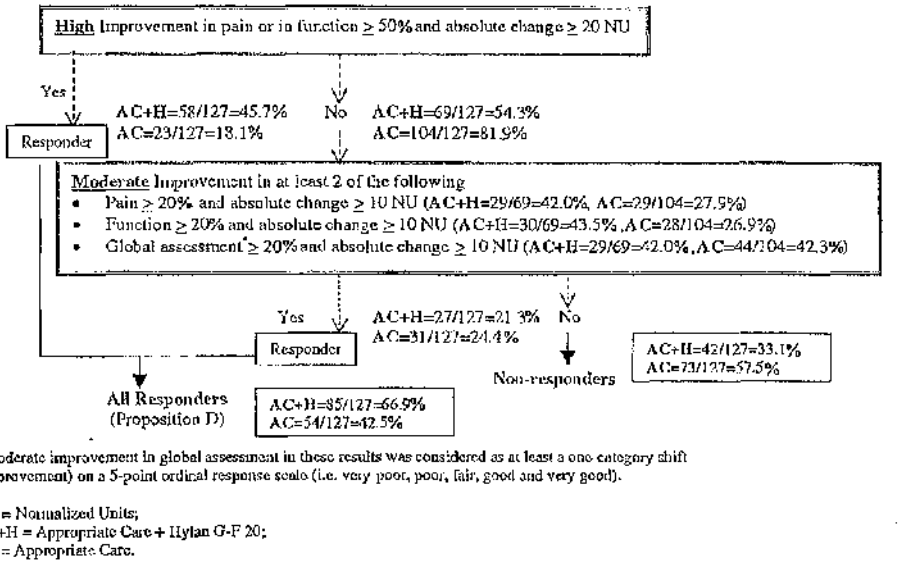


Fig. 3. Percentage of patients fulfilling specific criteria and overall response criteria for Proposition D.

STATISTICAL METHODS

Data from the locked study database were analyzed using SAS version 8.2. A logistic analysis was performed using the new OARSI responder criteria defined above (Propositions A, B, and D) with treatment group. A patient was classified as a "responder" if the patient improved according to the criteria outlined in the proposition from baseline to month 12. The hypotheses to be tested were whether AC+H was superior to AC when each of the new OARSI responder criteria were applied, and if treatment group differences were significantly different from each other.

All analyses were adjusted for design variables (i.e., baseline value of the variable being analyzed, site, blocking by site, body mass index, and baseline WOMAC total score). The type 1 experiment-wise error rate was controlled for by distributing alpha over all three comparisons (i.e., Propositions A, B, and D) using the Bonferroni adjustment of alpha/3 (alpha for each comparison = 0.05/3 = 0.0166). The original secondary effectiveness measure (i.e., percentage of patients with at least 20% improvement in the study knee at termination since baseline) was analyzed using a logistic analysis that controlled for the above-mentioned design variables. In this analysis, the type 1 experiment-wise error rate was not adjusted as there was one secondary effectiveness measure. The results of the original secondary analysis are provided for comparison to the new OARSI responder criteria.

All patients were included in the intent-to-treat group for all analyses as described earlier[1]. The hot deck method was used to impute data for all effectiveness measures as described earlier[1].

Results

PATIENTS

In the trial, 128 patients were randomized to receive appropriate care and 127 patients were randomized to receive appropriate care with hylan G-F 20. A total of 24 patients dropped out of the study (21 in the AC group, 3 in the AC+H group). One patient in the AC group did not have a baseline WOMAC questionnaire completed and was thus not included in the analysis. Descriptive statistics comparing demographics, baseline disease characteristics and baseline outcome measures (i.e., WOMAC, global assessment) are provided in Tables I and II. Overall, treatment groups were similar with respect to demographics, disease characteristics, and OA therapies used at baseline. However, 20% of patients in the AC+H group and 33% of patients in the AC group had grade IV OA, as subsequently determined by central radiologic grading. In addition, 18% of patients in the AC+H group and 30% of patients in the AC group had a 'very poor' global assessment of OA in the study knee, at baseline. Overall WOMAC scores for pain, stiffness and function were similar between groups.

KNEE OA TREATMENT

All patients except one in the AC+H group had at least one course of hylan G-F 20 in their study knee, and 53 (42%) had at least one course in their contralateral knee. Forty-five patients (38%) in the AC+H group received a second course of hylan G-F 20 in their study knee, and three received a third course in their study knee. Twenty patients (16%) in the AC+H group received a second course in their contralateral knee. There were more patients in the AC vs the AC+H group who reported corticosteroid injection(s) in the study knee (89 vs 18) or the other knee (35 vs 8) (both $P < 0.0001$). There were also more patients in the AC group taking NSAIDs for any knee (101 vs 82) ($P = 0.0062$), and other medications (25 vs 13) ($P = 0.0216$) (e.g., opioid analgesics, anti-inflammatories) for any knee. There was no significant difference between the groups in the utilization of concomitant medications for overall OA. Further details of knee OA treatment can be found in the original study results[1].

EFFECTIVENESS

The overall results for the original secondary effectiveness measures and the new OARSI responder criteria measures are provided in Table III. Figures 1–3 provide

Table I
Demographics and disease status at baseline

Baseline measure	AC+H, $n = 127$	AC, $n = 128$
	Mean (SD)	Mean (SD)
Age (years)	62.6 (9.4)	63.5 (10.5)
Body mass index (kg/m^2)	32.1 (8.0)	32.9 (7.2)
Duration (years) of OA symptoms		
Study knee	9.0 (9.5)	9.9 (9.7)
Other knee	7.4 (8.8)	8.3 (9.3)
	n (%)	*n* (%)
Sex: female	86 (68)	93 (73)
Previous therapy for OA of the knee(s)		
Acetaminophen	100 (79)	109 (85)
NSAIDs	120 (94)	110 (86)
Prior surgery, study knee	40 (31)	39 (30)
Prior surgery, other knee	40 (31)	23 (18)
Radiology grading within 1 year (central grading*)		
Not reported	0 (0)	1 (1)
Grade 0	4 (3)	4 (3)
Grade I	17 (13)	11 (9)
Grade II	32 (25)	33 (26)
Grade III	49 (39)	37 (29)
Grade IV	25 (20)	42 (33)
OA at baseline		
Other knee affected	109 (86)	108 (84)
Other joints affected	95 (75)	87 (68)

*Radiology grading is based on central grading, which may have differed from the site investigator's determination for patient eligibility.

more detail by describing the specific percentage of patients fulfilling the separate criteria within Propositions A, B, and D. The results show that for all definitions of improvement, the percentage of responders was greater in the AC+H group vs the AC group. Also, the differences were statistically significant at the 0.0166 level ($\alpha/3 = 0.05/3$) for all definitions of patient improved (adjusted using the Bonferroni correction) (Table III). The incremental difference between treatment groups always exceeded the required 20% difference established *a priori* as the minimum clinically important difference between groups. The magnitude of the treatment differences was similar across all definitions of improvement, varying from 24.4% (Propositions B and D) to 28.3% (Proposition A, and '20% improvement in pain' criteria).

Table II
WOMAC and patient global assessment at baseline

Baseline measure	AC+H, $n = 127$	AC, $n = 128$
	Mean (SD)	Mean (SD)
WOMAC pain (0–20)*	11.4 (2.7)	11.9 (2.9)
WOMAC stiffness (0–8)*	5.1 (1.5)	5.1 (1.4)
WOMAC function (0–68)*	39.5 (9.3)	40.2 (9.3)
Patient global assessment of OA in study knee	*n* (%)	*n* (%)
Not reported	0 (0)	1 (1)
Very good	0 (0)	0 (0)
Good	2 (2)	1 (1)
Fair	44 (35)	31 (24)
Poor	58 (46)	57 (45)
Very poor	23 (18)	38 (30)

*The higher the score, the worse the problem. One patient in the AC group did not have a WOMAC questionnaire completed at baseline and was thus not included in the analysis.

When comparing the new OARSI responder criteria, the highest effectiveness in both treatment groups was observed with Proposition D (66.9% vs 42.5%). Conversely, Proposition A resulted in the lowest effectiveness in both treatment groups (53.5% vs 25.2%). The difference between treatment groups was similar for all propositions (24.4%–28.3%) and was generally similar for the original secondary measures of effectiveness.

Figures 1–3 describe the percentage of patients fulfilling the specific criteria in each of Propositions A, B, and D. In the AC+H group, two thirds of the patients fulfilled the criteria of a 'responder' by having a 'high' improvement in pain (Proposition A) or 'high' improvement in pain or function (Propositions B and D). For example, for Proposition D, 45.7% of the AC+H group were considered responders since they had a 'high' improvement response in pain or function, where the overall percentage of responders was 66.9% (i.e., 45.7%/66.9% = 68.3% or two thirds). For the patients in the AC group, two thirds of patients (i.e., 16.5%/25.2% = 65.5% or two thirds) in Proposition A fulfilled the criteria of responder by having a 'high' improvement in pain, however, less than half of patients in Propositions B and D fulfilled the criteria of responder by having had 'high' improvement in pain or function.

As the cut-off criteria for 'moderate' improvement domains (pain, function, global assessment) become more stringent moving from Propositions D to B to A, the percentage of responders (i.e., by moderate improvement only) in both treatment groups decreases (AC+H: 21.3%, 17.3%, 16.5%; AC: 24.4%, 18.1%, 8.7%, respectively). The results also show that for Propositions A and B, the greatest percentage of patients fulfilling 'moderate' improvement criteria are numerically with the global assessment domain (AC+H: 48.8%, 45.4%; AC: 44.3%, 43.1%, respectively). For Proposition D, the percentage of patients fulfilling 'moderate' improvement criteria was similar across the domains of pain and function and global assessment for the AC+H group, but was higher for global assessment for the AC group.

Discussion

Traditional methods of performing between-group comparisons of clinical trials data are frequently based on the analysis of continuous variables. These provide an appreciation of the magnitude and variation in group effects, but do not permit reviewers an understanding of the degree of improvement experienced by individual patients. In contrast, responder criteria, while being reductionist from a group standpoint, are capable of categorizing individual patients according to whether they achieve levels of improvement at or above pre-specified response thresholds. These thresholds have been established *a priori* either to reflect a clinically important difference at an individual level, or on the basis of differentiating most efficiently between an active treatment and a placebo control. In the case of the effectiveness measures used in the original study[1], they were proposed during protocol development at a time when there was no precedent to follow, but 20% was considered, by the development group, to represent a minimum clinically important difference. This difference was one that was of the same order of magnitude as previously published ACR 20 criteria[15] for rheumatoid arthritis clinical trials. The OARSI responder criteria[7] were developed during the execution of the protocol, and the OMERACT–OARSI responder criteria[8] were developed following completion of the study. It is

Table III
Summary of response criteria: improvement from baseline to month 12

Criteria	Treatment group		Difference	
	AC+H, $n = 127$ n (%)	AC*, $n = 127$ n (%)	[(AC+H) − (AC)] (%)	P-value
Original effectiveness measures				
≥ 20% improvement in pain†	87 (68.5)	51 (40.2)	28.3	0.0001
≥ 20% improvement in pain and either ≥ 20% improvement in function or stiffness†	79 (62.2)	45 (35.4)	26.8	0.0001
New OARSI response criteria				
Proposition A (OARSI)	68 (53.5)	32 (25.2)	28.3	<0.0001
Proposition B (OARSI)	72 (56.7)	41 (32.3)	24.4	0.0002
Proposition D (OMERACT–OARSI)	85 (66.9)	54 (42.5)	24.4	0.0002

*One patient in the AC group did not have a WOMAC questionnaire completed at baseline and was thus not included in the analysis.
†Secondary effectiveness measures from main study.

of interest therefore that these different approaches, separated in time and concept, all yield comparable, statistically detectable between-group differences, in favor of AC+H, of 24.4%–28.3%.

It is worthy of note that different criteria, while yielding very similar between-group differences (24.4%–28.3%) (Table III), specify quite different percentages of both AC+H (53.5%–68.5%) and AC alone (25.2%–42.5%) patients as responders. For this reason, the criteria should not be considered interchangeable, but rather complementary. The OARSI–OMERACT Proposition D, and the WOMAC 20% improvement in pain response criteria, yield very similar by-group percentage response rates and similar between-group differences in this study. The generalizability of this observation will require further study. This similarity in results may be explained by the requirement for those not fulfilling the primary requirement in Proposition D, of alternatively experiencing an improvement of 20% or greater in pain, function or global assessment. In contrast, Propositions A and B require secondary improvements of 30% (Proposition A) or 20% or greater, respectively (Proposition B). Thus, the similarity between the lowest threshold for pain on Proposition D and WOMAC 20% improvement in pain criteria are similar in percentage terms, and may explain the comparable yields. However, since there is no absolute change required for the WOMAC 20% improvement in pain criteria, whereas in Proposition D an improvement of at least 10 NU is required, the two approaches may yield different results with differing response distributions, and may therefore differ from study to study.

Furthermore, the patients in this study had mean baseline scores in or about the midpoint on all three WOMAC subscales. Whether these observations can be generalized to patients with either more or less severe symptoms will require further study. The pattern of improvement in patient global assessment is different from that in pain and function. For Propositions A and B, the percentage of responders for moderate improvement in global assessment is greater than the percentage of responders for moderate improvement in pain or function (Figs. 1, 2). Hoeksma *et al.*[16] have observed that fewer patients with hip OA were classified as improved with OARSI response criteria as compared to patient's global assessment. This suggested to the investigators that the OARSI response criteria provided a "more objective reflection of the actual clinical status of patients with OA of the hip". Taken collectively, these observations suggest that the wording and scaling of the patient global assessment question might require further consideration for application in studies employing responder analyses.

The results of this analysis provide evidence for the capacity of responder criteria to detect clinically important statistically detectable differences between two active treatment groups in a randomized clinical trial. These observations support those of other investigators who have reported their experience on the responsiveness of the OARSI responder criteria[17–19]. In particular, this approach allows reviewers and consumers to discern how many patients experienced a clinically important reduction in the severity of their symptoms. Given that the analytic strategy is individualized, this approach may have important implications for monitoring patients in routine clinical care and facilitating evidence-based therapeutic decision making, and shared goal setting, in various health care environments.

Acknowledgment

The authors thank Genzyme Corporation for study funding. The study agreement with Genzyme Corporation gave the investigators independence to publish regardless of the results. The authors would like to thank the clinical investigators who enrolled patients and are listed in the main manuscripts[1,11].

References

1. Raynauld JP, Torrance GW, Band PA, Goldsmith CH, Tugwell P, Walker V, *et al.* A prospective, randomized, pragmatic, health outcomes trial evaluating the incorporation of hylan G-F 20 into the treatment paradigm for patients with knee osteoarthritis (part 1 of 2): clinical results. Osteoarthritis Cartilage 2002;10: 506–17.
2. Bellamy N. Musculoskeletal Clinical Metrology. Lancaster: Kluwer Academic 1993.
3. Bellamy N, Carr A, Dougados M, Shea B, Wells G. Towards a definition of "difference" in osteoarthritis. J Rheumatol 2001;28:427–30.
4. Bellamy N, Lybrand SG, Gee T. An evaluation of the convergence between three different methods of response status assignment (RSA) based on the WOMAC osteoarthritis index (Abstract). Arthritis Rheum 2002;46(Suppl) 339:pS159.

5. Ehrich EW, Davies GM, Watson DJ, Bolognese JA, Seidenberg BC, Bellamy N, *et al.* Minimal perceptible clinical improvement with the Western Ontario and McMaster Universities Osteoarthritis Index questionnaire and global assessments in patients with osteoarthritis. J Rheumatol 2000;27:2635–41.
6. Beaton DE, Bombardier C, Katz JN, Wright JG, Wells G, Boers M, *et al.* Looking for important change/difference in studies of responsiveness. J Rheumatol 2001;28:400–5.
7. Dougados M, LeClaire P, van der Heijde D, Bloch DA, Bellamy N, Altman RD. Response criteria for clinical trials on osteoarthritis of the knee and hip: a report of the Osteoarthritis Research Society International standing committee for clinical trials response criteria initiative. Osteoarthritis Cartilage 2000;8:395–406.
8. Pham T, van der Heijde D, Lassere M, Altman RD, Anderson JJ, Bellamy N, *et al.* Outcome variables for osteoarthritis clinical trials: the OMERACT–OARSI set of responder criteria. J Rheumatol 2003;30:1648–54.
9. Bellamy N, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P, *et al.* Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip and hand osteoarthritis. Consensus development at OMERACT III. J Rheumatol 1997;24: 799–802.
10. Osteoarthritis Research Society (ORS) Task Force Report: design and conduct of clinical trials of patients with osteoarthritis: recommendations of a task force of the Osteoarthritis Research Society. Osteoarthritis Cartilage 1996;4:217–43.
11. Torrance G, Raynaud JP, Walker V, Goldsmith CH, Bellamy N, Band P, *et al.* A prospective randomized health outcomes trial evaluating the incorporation of hylan G-F 20 into the treatment paradigm for patients with knee osteoarthritis (part 2 of 2): economic results. Osteoarthritis Cartilage 2002;10:518–27.
12. Hochberg MC, Altman RD, Brandt KD, Clark BM, Dieppe PA, Griffin MR, *et al.* Guidelines for the medical management of osteoarthritis. Part II. Osteoarthritis of the knee. Arthritis Rheum 1995;38:1541–6.
13. Bellamy N, Buchanan WW. Clinical evalution in the rheumatic diseases. In: Arthritis and Allied conditions, Koopman WJ, Moreland LK, Eds. London: Lippincott, Williams and Wilkins, 15th Edition 2005; 51–80.
14. Goldsmith CH, Boers M, Bombardier C, Tugwell P. Criteria for clinically important changes in outcomes: development, scoring and evaluation of rheumatoid arthritis patient and trial profiles. OMERACT Committee. J Rheumatol 1993;20:561–5.
15. Felson DT, Anderson JJ, Boers M, Bombardier C, Furst D, Goldsmith CH, *et al.* American College of Rheumatology preliminary definition of improvement in rheumatoid arthritis. Arthritis Rheum 1995;38:727–35.
16. Hoeksma HL, van den Ende E, Ronday HK. A comparison of the OARSI response criteria with patient's global assessment in patients with osteoarthritis of the hip treated with manual therapy or exercise therapy (Abstract). Osteoarthritis Cartilage 2003;11(Suppl A):S36.
17. Verges J, Montell E, Martinez G, Coronas C, Du Souich P. Chondroitin sulphate (CS) complies with the therapeutic response criteria recently proposed by the OARSI for knee osteoarthritis (Abstract). Ann Rheum Dis 2003; 62(Suppl 1):AB0442.
18. Yocum DE, McGoldrick K, Waller PR, Segarra R, Páez A, Pavesi M, *et al.* LAS 34475, a novel Cox-2 inhibitor is effective in the treatment of osteoarthritis of the knee according to the OARSI criteria (Abstract). Ann Rheum Dis 2003;62(Suppl 1):AB0429.
19. Gagnier P, Hwang E. The use of the OARSI responder index as a tool to assess dose response relationships in OA efficacy trials (Abstract). Ann Rheum Dis 2002; 61(Suppl 1):133.

EXTENDED REPORT

Evaluation of clinically relevant changes in patient reported outcomes in knee and hip osteoarthritis: the minimal clinically important improvement

F Tubach, P Ravaud, G Baron, B Falissard, I Logeart, N Bellamy, C Bombardier, D Felson, M Hochberg, D van der Heijde, M Dougados

Ann Rheum Dis 2005;64:29–33. doi: 10.1136/ard.2004.022905

See end of article for authors' affiliations

Correspondence to: Dr F Tubach, Département d'Epidémiologie, Biostatistique et Recherche Clinique, INSERM E0357, Hôpital Bichat, 46 rue Henri Huchard, 75018 Paris, France; florence.tubach@bch.ap-hop-paris.fr

Accepted 9 May 2004
Published Online First 18 June 2004

Background: In clinical trials, at the group level, results are usually reported as mean and standard deviation of the change in score, which is not meaningful for most readers.
Objective: To determine the minimal clinically important improvement (MCII) of pain, patient's global assessment of disease activity, and functional impairment in patients with knee and hip osteoarthritis (OA).
Methods: A prospective multicentre 4 week cohort study involving 1362 outpatients with knee or hip OA was carried out. Data on assessment of pain and patient's global assessment, measured on visual analogue scales, and functional impairment, measured on the Western Ontario McMaster Universities Osteoarthritis Index (WOMAC) function subscale, were collected at baseline and final visits. Patients assessed their response to treatment on a five point Likert scale at the final visit. An anchoring method based on the patient's opinion was used. The MCII was estimated in a subgroup of 814 patients (603 with knee OA, 211 with hip OA).
Results: For knee and hip OA, MCII for absolute (and relative) changes were, respectively, (*a*) −19.9 mm (−40.8%) and −15.3 mm (−32.0%) for pain; (*b*) −18.3 mm (−39.0%) and −15.2 mm (−32.6%) for patient's global assessment; (*c*) −9.1 (−26.0%) and −7.9 (−21.1%) for WOMAC function subscale score. The MCII is affected by the initial degree of severity of the symptoms but not by age, disease duration, or sex.
Conclusion: Using criteria such as MCII in clinical trials would provide meaningful information which would help in interpreting the results by expressing them as a proportion of improved patients.

The choice of an outcome measure is a major step in the design of clinical trials. In evaluating the symptomatic severity of osteoarthritis (OA) of the lower limbs, scientific groups such as the OMERACT (Outcome Measures in Rheumatology Group),[1] GREES (Group for the Respect of Ethics and Excellence in Science),[2] and OARSI (OsteoArthritis Research Society International)[3] have raised the importance of evaluating at least three dimensions: pain, patient's global assessment of disease status, and functional impairment. At the individual level, determining the minimal meaningful change in a score by use of a structured instrument is a challenge. Are changes in self reported levels of pain of 10 mm on a 0–100 mm visual analogue scale (VAS) clinically important? Does the change reflect meaningful improvement for the patient? The concept of the minimal clinically important difference (MCID)[4–6] could help in interpreting changes in scores at the individual level. However, the MCID, which can reflect either an improvement or a worsening, has not been used here, because in clinical trials we are always interested in improvement and not worsening. Furthermore, it has been shown that the MCID could be different for improvement and worsening.[7] The minimal clinically important improvement (MCII), defined as the smallest change in measurement that signifies an important improvement in a patient's symptom, seems more appropriate and, in clinical trials, provides readers with additional information on the effect size by expressing the results more meaningfully (that is, as a percentage of improved patients).

This prospective cohort study aimed at estimating the MCII from the patient's perspective for three main patient reported outcomes used in OA trials: pain, patient's global assessment of disease activity, and functional impairment.

MATERIALS AND METHODS

Study design

We conducted a prospective 4 week cohort study.

Study population

This study involved 1362 outpatients with knee or hip OA, as defined by the American College of Rheumatology,[8 9] included by 399 rheumatologists. Each rheumatologist had to recruit four patients, three with knee OA and one with hip OA. To be included in the study, patients had to experience pain from OA (≥30 mm on a VAS varying from 0 to 100), require treatment with a non-steroidal anti-inflammatory drug (NSAID), and be able to complete questionnaires in French. Inclusion could begin with the onset of treatment or a switch from one NSAID to another. Patients were excluded if they had a prosthesis on the assessed joint or if they had been given an intra-articular injection in the 4 weeks before the study began. All patients initially visited the rheumatologist in charge of their case, and an NSAID was prescribed (the drug and its dosage was chosen by the physician). A final visit to the same rheumatologist was scheduled 4 weeks later.

Abbreviations: MCID, minimal clinically important difference; MCII, minimal clinically important improvement; NSAID, non-steroidal anti-inflammatory drug; OA, osteoarthritis; VAS, visual analogue scale; WOMAC, Western Ontario McMaster Universities Osteoarthritis Index

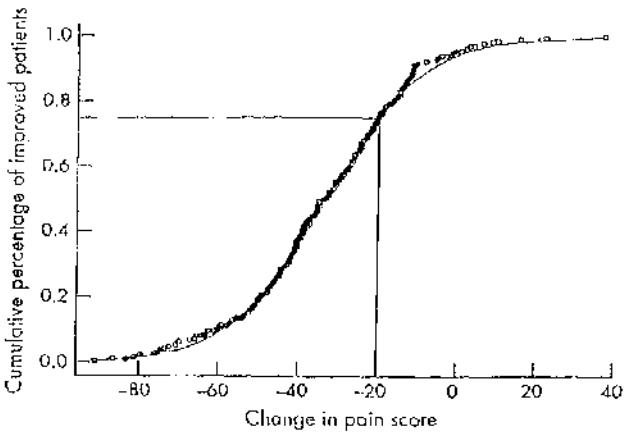


Figure 1 Aspects of the cumulative distribution function used to determine the MCII (changes in pain score in patients with knee OA; n = 265). Among patients considering their response to treatment as good on a five point Likert scale, 75% experienced a decrease in pain between baseline and final visit of ≥19.9 mm on a 0–100 mm VAS (a change between −100 mm and −19.9 mm).

Measurements

At the baseline visit, demographic and disease data were collected. Patients assessed their OA status at baseline and final visit. They assessed the following patient reported outcomes: (*a*) pain on movement during the 48 hours before the visit, measured on a 0–100 mm VAS; (*b*) global assessment of disease activity measured on a 0–100 mm VAS; and (*c*) physical function, measured on the Western Ontario McMaster Universities Osteoarthritis Index (WOMAC) function subscale (17 items, five point Likert scale for each item; high scores indicate high degree of functional impairment; total score normalised to a 0–100 score).

At the final visit, a random sample of two thirds of the patients (n = 923) assessed their response to NSAID treatment on a five point Likert scale (none = no good at all, ineffective drug; poor = some effect but unsatisfactory; fair = reasonable effect but could be better; good = satisfactory effect with occasional episodes of pain or stiffness; excellent = ideal response, virtually pain free). The other third of the patients assessed their response to treatment on a 15 point Likert scale (from −7, a very great deal worse, to +7, a very great deal better, with 0, no change).

Statistical analysis

All the analyses considered patients with knee and hip OA separately.

The MCII was determined in a subgroup of 814 patients (603 with knee and 211 with hip OA) whose assessment of response to treatment was measured on a five point Likert scale and who had completed the final visit.

An anchoring method based on the patient's assessment of response to treatment was used.

The MCII was estimated for both the absolute (final value−baseline value) and the relative ((final value−baseline value)/baseline value) changes in each patient reported outcome. It was estimated by constructing a curve of cumulative percentages of patients as a function of the change in score (for example, difference in pain score) among patients whose final evaluation of response to treatment was "good, satisfactory effect with occasional episodes of pain or stiffness", because we wanted to focus on the improvement that was clinically important. Logistic regression was used to model the observations (fig 1). We targeted the point at the flattening of the curve at which most subjects stated they had improved. To determine the change in score corresponding to this point, we first looked at the two parameter logistic model that best fitted the data. Then we determined the square root of the third derivative of this logistic function that corresponded with the MCII. One can demonstrate that this point corresponds by construction to the 78.9th centile of the change in score, and thus we propose to define the MCII as the 75th centile of the change in score, because it is very close to the point defined above and easier to derive. The model permitted us, firstly, to determine that the target point was

Table 1 Baseline characteristics of patients

Characteristics	Knee OA (n=603)		Hip OA (n=211)	
	Mean	SD	Mean	SD
Age (years)	67.9	10.2	64.6	10.2
Weight (kg)	75.5	13.8	71.3	12.3
Height (cm)	163.8	8.5	164.7	8.4
Body mass index (kg/m²)	28.1	4.8	26.2	3.8
Disease duration (years)	4.7	5.8	3.3	4.6
Pain score (0–100 mm VAS)				
Week 0	59.3	16.2	56.7	16.5
Change (week 0–week 4)	−24.9	21.5	−20.0	21.7
Patient global assessment (0–100 mm VAS)				
Week 0	59.6	18.3	58.0	19.3
Change (week 0–week 4)	−24.7	24.0	−20.6	23.2
WOMAC function score (0–100)				
Week 0	42.8	16.1	44.4	16.5
Change (week 0–week 4)	−11.6	13.9	−10.4	13.6
	No	%	No	%
Female sex	421	69.8	133	63.0
Kellgren & Lawrence grade				
II	108	17.9	33	15.7
III	268	44.4	111	52.9
IV	227	37.7	66	31.4
NSAID* intake during past 4 weeks	178	29.7	69	32.7
Analgesic treatment†	344	57.2	141	67.1
Symptomatic slow-acting drug intake‡	209	34.8	90	42.9

*Non-steroidal anti-inflammatory drugs (before the start of the study); †other than NSAIDs (before the start of the study); ‡chondroitin sulphate, diacerhein, or avocado/soybean unsaponifiables.

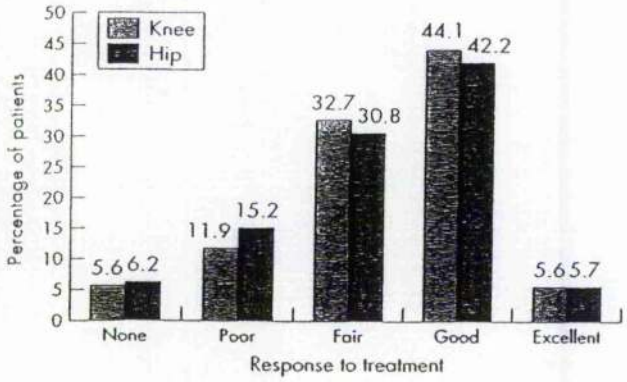


Figure 2 Patients' assessment of their response to treatment.

correctly approached by the 75th centile and, secondly, to estimate the 95% confidence intervals.

In a second step, we stratified the analysis on the baseline score of interest (divided into tertiles) to assess whether the level of pain, the patient's assessment of disease activity, and functional impairment had a modifying effect on the MCII. That is we stratified (*a*) on the baseline pain score to estimate the MCII for pain; (*b*) on the baseline assessment of disease activity to estimate the MCII for patient's assessment of disease activity; (*c*) on the baseline WOMAC function score to estimate the MCII for functional impairment.

In a third step, to investigate the effect of covariates (other than location of OA) on the MCII, we stratified the analysis successively by age, disease duration (both divided into tertiles), and sex.

Statistical analyses was performed with the SAS Release 8.2 statistical software package and the S plus 4.5 statistical software package.

Compliance with research ethics standards

This study was conducted in compliance, with the protocol, good clinical practices, and the Declaration of Helsinki principles.

RESULTS

A total of 1362 patients were enrolled in the study: 1019 (75%) had knee and 343 (25%) hip OA; 913 (67%) were female; and the mean (SD) age was 67.2 (10.5) years. A total of 914 (90%) patients with knee and 310 (90%) with hip OA completed the final visit. Patients lost to follow up were excluded from the analysis and did not differ from completers in their baseline characteristics. Among the completers, 603 patients with knee and 211 with hip OA assessed their response to treatment on a five point Likert scale.

Table 1 shows the descriptive statistics on clinical and demographics variables. Figure 2 shows patients' rating of response to treatment.

Table 2 lists the MCII values for the three patient reported outcomes, according to location of OA. These values were estimated in the 265 patients with knee and the 87 patients with hip OA who completed the final visit and assessed their response to treatment as "good". For instance, patients with knee OA considered themselves clinically improved if the decrease in pain exceeded 19.9 mm on the 0–100 mm VAS. We used the data from the five point not the 15 point Likert scale mentioned in the "Methods" section.

Table 3 shows the estimates of the MCII (for absolute change) stratified on the baseline score in patients with knee or hip OA. The higher the baseline score, the larger the MCII. Patients who have a severe symptom need a higher level of change to consider themselves clinically improved than those with less severe symptoms. For instance, patients with severe pain (a high tertile of baseline pain score) considered themselves clinically improved if the decrease in pain exceeded 36.6 mm on the 0–100 mm VAS. Patients with less pain (low tertile of baseline pain score) needed a lower level of change (−10.8 mm on the VAS) to consider themselves clinically improved. The estimates of the MCII for relative change also varied across tertiles of the baseline score (data not shown).

The estimates of the MCII do not vary across age, disease duration tertiles, or sex (data not shown).

DISCUSSION

This study dealt with the clinical meaningfulness of changes observed for patient reported outcome measures. Because a statistically significant difference is mostly a matter of sample size, the most difficult issue is whether an observed or estimated difference is clinically important.[10] In other words, statistical significance is not equivalent to clinical significance. Reporting results of a trial using the MCII (that is, as a percentage of improved patients) provides readers with values which are more easily understood and additional information to help them decide whether a treatment should be used. This threshold also allows for monitoring of individual response to treatment over time and adapting treatment to individual patients (for example, determining whether to start or interrupt a treatment). Furthermore, the designation and use of MCII in clinical trials is critical for meaningful systematic reviews and combining results from different studies in meta-analyses.[11] This concept aims at complementing, not replacing, information on the effect size, because the effect size remains a more powerful approach.[12]

The MCII is the smallest change in measures that signifies an important improvement in a patient's symptom. Thus, the MCII can undoubtedly be considered as a treatment target from the patient's perspective. It is based on the patient's opinion as an external anchor and contrasts changes within

Table 2 Minimal clinically important improvement (MCII) scores according to patients' location of OA

	Knee OA				Hip OA			
	Absolute change		Relative change		Absolute change		Relative change	
Patient reported outcomes	MCII	(95% CI)	MCII (%)	(95% CI)	MCII	(95% CI)	MCII (%)	(95% CI)
Pain (0–100 mm VAS), mm	−19.9	(−21.6 to −17.9)	−40.8	(−44.8 to −36.1)	−15.3	(−17.8 to −12.5)	−32.0	(−38.5 to −24.0)
Patient global assessment (0–100 mm VAS), mm	−18.3	(−19.8 to −16.7)	−39.0	(−45.8 to −30.6)	−15.2	(−16.9 to −13.4)	−32.6	(−38.7 to −25.2)
WOMAC function score (0–100)	−9.1	(−10.5 to −7.5)	−26.0	(−28.6 to −23.3)	−7.9	(−8.8 to −5.0)	−21.1	(−24.8 to −17.0)

The MCII was defined as the 75th centile of the change in score among patients whose evaluation of response to treatment was "good", for the three patient reported outcomes: pain, as assessed on a visual analogue scale (VAS); global assessment of disease status, on a VAS; or the Western Ontario McMaster Universities Osteoarthritis Index (WOMAC) function subscale score.

Table 3 Minimal clinically important improvement (MCII) score of absolute change in patients with knee or hip OA, by low, intermediate, and high baseline score tertiles

	Knee OA			Hip OA		
	Baseline score tertile			Baseline score tertile		
	Low	Intermediate	High	Low	Intermediate	High
Pain (0–100 mm VAS)	−10.8 (−12.7 to −8.7) (30 to 51.0)	−27.4 (−29.7 to −24.6) (51.1 to 66.2)	−36.6 (−38.3 to −34.7) (>66.2)	−7.2 (−10.7 to −2.9) (30 to 49.3)	−23.9 (−28.3 to −18.0) (49.4 to 65.4)	29.7 (−35.4 to −21.8) (>65.4)
Patient's global assessment of disease (0–100 mm VAS)	−6.4 (−8.6 to −3.8) (≤50.4)	−24.6 (−26.8 to −22.1) (50.5 to 68.7)	−43.2 (−46.8 to −38.7) (>68.7)	−4.3 (−6.9 to −1.4) (≤49.9)	−26.0 (−28.3 to −23.3) (50.0 to 69.9)	−29.9 (−34.5 to −24.3) (>69.9)
WOMAC function score (0–100)	−5.3 (−6.5 to 3.8) (≤35.3)	−11.8 (−13.0 to −10.4) (35.4 to 51.5)	−20.4 (−22.5 to −18.1) (>51.5)	2.6 (−4.4 to −0.5) (≤38.2)	−14.8 (−17.0 to −12.0) (38.3 to 52.9)	−15.1 (−18.9 to −10.0) (>52.9)

Results are shown as mean (95% confidence interval) (tertile).
The MCII was defined as the 75th centile of the change in score among patients whose evaluation of response to treatment was "good", for three patient reported outcomes: pain, as assessed on a visual analogue scale (VAS); global assessment of disease status, on a VAS; or the Western Ontario McMaster Universities Osteoarthritis Index (WOMAC) function subscale score.

patients at the individual level (proportion of improved patients) instead of at the group level (mean change in a variable).

Approaches such as investigator defined (expert consensus) or statistically defined methods have been used to determine this threshold.[13 14] Despite the absence of a criterion measure, establishing the meaning of changes in a measure requires an independent standard. Patient global ratings are recommended as an external anchor for evaluating the clinical significance of individual change.[13] The large sample of patients as experts in determining improvement is a good indicator of representativeness.

To determine the MCII, the external criterion was the patient's assessment of response to treatment as assessed on a five point Likert scale. We defined MCII in the group of patients whose evaluation of response to treatment was "good", because one is always looking for clinically important differences. We did not include patients whose evaluation of response to treatment was "excellent," because our target was the *minimal* change important from the patient's perspective. But obviously, this choice was arbitrary and affects the results (data not shown). The group of patients in whom MCII is determined and the wording of the items in the questionnaire to assess response to treatment should be chosen with the help of experts; in our study, the group of patients were chosen by the experts NB, CB, DF, MH, DvdH, and MD.

In a previous study,[15] a three round Delphi method involving six academic rheumatologists experienced in OA trials was used to define the MCID for some outcome measures used in OA trials (not specifically focusing on hip or knee OA). The MCID for patient pain on movement (measured on a 0–100 VAS) was 17.5 mm and that for patient global assessment of disease activity (measured on a 0–100 VAS) was 15 mm. Although this method differs from that used in our study, the values are very close to our estimates of MCII for these patient reported outcomes. The only study dealing with meaningful change for the WOMAC dealt with the minimal clinically perceptible difference not the MCID.[16]

Our study has demonstrated that the MCII varies depending on the baseline state. Patients who have the most severe symptoms have to experience a greater change to consider themselves improved. Riddle and colleagues also found this effect in their investigation of low back pain,[17] where the MCID varied between 3 and 13 depending on the baseline range of scores (on the Roland-Morris Back Pain Questionnaire,[18] total score varying from 0 to 24 points, with baseline scores divided into five approximately equal sized intervals). However, the precision of their estimates may have been compromised by the small sample size, especially for patients with high levels of disability.

The variation of MCII across tertiles of baseline scores in our study cannot been imputed to the size of the sample, as confirmed by the narrowness of the 95% confidence intervals. We believe that this variation depending on the baseline score may preclude the use of the crude MCII. The patient's initial or previous score should be taken into account when making decisions about important change. We propose to use three estimates of MCII, corresponding to the tertiles of each baseline score, to express the changes in terms of important improvement. This meets the recommendation of Crosby and associates[13] for estimating MCID in health related quality of life criteria: to anchor baseline severity of individual patients.

We believe this is the first study to investigate the effect of several covariates such as age, sex, OA location, and disease duration on patient responses. It is interesting to observe that these factors do not consistently modify the estimates of MCII.

In conclusion, use of the concept MCII facilitates the presentation and interpretation of results obtained in clinical trials and the transposition of trial results into practice. However, the baseline score should be taken into account. Further studies involving different datasets, clinical environments, languages, and countries are necessary to validate these observations prospectively.

ACKNOWLEDGEMENTS

This study was supported by an unrestricted grant from Merck, Sharp, & Dohme Chibret Laboratories, France.

Authors' affiliations

F Tubach, P Ravaud, G Baron, Institut National de la Santé et de la Recherche Médicale (INSERM) E 0357; Département d'Epidémiologie, Biostatistique et Recherche Clinique; Groupe Hospitalier Bichat-Claude Bernard (Assistance Publique—Hôpitaux de Paris); Faculté Xavier Bichat (Université Paris 7), Paris, France
B Falissard, Faculté de Médecine Paris-sud, Département de Santé Publique, Hôpital Paul Brousse (Assistance Publique—Hôpitaux de Paris), Villejuif, France
I Logeart, Merck, Sharp & Dohme Chibret Laboratories, Paris, France
N Bellamy, Department of Medicine, University of Queensland, Royal Brisbane Hospital, Brisbane, Queensland, Australia
C Bombardier, Institute for Work and Health, Toronto, Ontario, Canada
D Felson, Boston University School of Medicine, Boston, Massachusetts, USA

M Hochberg, University of Maryland, Baltimore, Maryland, USA
D van der Heijde, University Hospital, Maastricht, The Netherlands
M Dougados, Service de Rhumatologie B, Hôpital Cochin (Assistance Publique—Hôpitaux de Paris), Paris, France

REFERENCES

1 **Bellamy N**, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P, *et al.* Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development of OMERACT III. *J Rheumatol* 1997;24:799–802.
2 **GREES.** Recommendations for the registration of drugs used in the treatment of osteoarthritis. Group for the respect of ethics and excellence in science (GREES): osteoarthritis section. *Ann Rheum Dis* 1996;55:552–7.
3 **Altman R**, Brandt K, Hochberg M, Moskowitz R, Bellamy N, Bloch DA, *et al.* Design and conduct of clinical trials in patients with osteoarthritis: recommendations from a task force of the Osteoarthritis Research Society. Results from a workshop. *Osteoarthritis Cartilage* 1996;4:217–43.
4 **Beaton DE**, Bombardier C, Katz JN, Wright JG. A taxonomy for responsiveness. *J Clin Epidemiol* 2001;54:1204–17.
5 **Juniper EF**, Guyatt GH, Willan A, Griffith LE. Determining a minimal important change in a disease-specific Quality of Life Questionnaire. *J Clin Epidemiol* 1994;47:81–7.
6 **Jaeschke R**, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989;10:407–15.
7 **Schwartz AL**, Meek PM, Nail LM, Fargo J, Lundquist M, Donofrio M, *et al.* Measurement of fatigue, determining minimally important clinical differences. *J Clin Epidemiol* 2002;55:239–44.
8 **Altman R**, Alarcon G, Appelrouth D, Bloch D, Borenstein D, Brandt K, *et al.* The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. *Arthritis Rheum* 1991;34:505–14.
9 **Altman R**, Asch E, Bloch D, Bole G, Borenstein D, Brandt K, *et al.* Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Association. *Arthritis Rheum* 1986;29:1039–49.
10 **Wright JG** Interpreting health-related quality of life scores: the simple rule of seven may not be so simple. *Med Care* 2003;41:597–8.
11 **Brooks P**, Tugwell P, Strand CV, Simon L International Consensus Conference on Outcome Measures in Rheumatology, Gold Coast, Australia, April 11–14, 2002. *J Rheumatol* 2003;30:866–7.
12 **Streiner DL**. Breaking up is hard to do: the heartbreak of dichotomizing continuous data. *Can J Psychiatry* 2002;47:262–6.
13 **Crosby RD**, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol* 2003;56:395–407.
14 **Beaton DE**. Simple as possible? Or too simple? Possible limits to the universality of the one half standard deviation. *Med Care* 2003;41:593–6.
15 **Bellamy N**, Corette S, Ford PM, Keun WT, le Riche NG, Lussier A, *et al.* Osteoarthritis antirheumatic drug trials. III. Setting the delta for clinical trials—results of a consensus development (Delphi) exercise. *J Rheumatol* 1992;19:451–7.
16 **Ehrich EW**, Davies GM, Watson DJ, Bolognese JA, Seidenberg BC, Bellamy N. Minimal perceptible clinical improvement with the Western Ontario and McMaster Universities osteoarthritis index questionnaire and global assessments in patients with osteoarthritis. *J Rheumatol* 2000;27:2635–41.
17 **Riddle DL**, Stratford PW, Binkley JM. Sensitivity to change of the Roland-Morris Back Pain Questionnaire: part 2. *Phys Ther* 1998;78:1197–207.
18 **Roland M**, Morris R. A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low-back pain. *Spine* 1983;8:141–4.

EXTENDED REPORT

Evaluation of clinically relevant states in patient reported outcomes in knee and hip osteoarthritis: the patient acceptable symptom state

F Tubach, P Ravaud, G Baron, B Falissard, I Logeart, N Bellamy, C Bombardier, D Felson, M Hochberg, D van der Heijde, M Dougados

Ann Rheum Dis 2005;64:34–37. doi: 10.1136/ard.2004.023028

See end of article for authors' affiliations

Correspondence to: Dr F Tubach, Département d'Epidémiologie, Biostatistique et Recherche Clinique, INSERM E0357, Hôpital Bichat, 46 rue Henri Huchard, 75018 Paris, France; florence.tubach@bch.ap-hop-paris.fr

Accepted 18 April 2004
Published Online First 18 June 2004

Background: The patient acceptable symptom state (PASS) is the value beyond which patients can consider themselves well. This concept can help in interpreting results of clinical trials.
Objective: To determine the PASS estimate for patients with knee and hip osteoarthritis (OA) by assessing pain, patient's global assessment of disease activity, and functional impairment.
Methods: A 4 week prospective multicentre cohort study of 1362 outpatients with knee or hip OA was carried out. Data on assessment of pain and patient's global assessment of disease, measured on visual analogue scales, and functional impairment, measured on the Western Ontario McMaster Universities Osteoarthritis Index (WOMAC) function subscale, were collected at baseline and final visits. The patients assessed their satisfaction with their current state at the final visit. An anchoring method based on the patient's opinion was used.
Results: For patients with knee and hip OA, the estimates of PASS were, respectively, 32.3 and 35.0 mm for pain, 32.0 and 34.6 mm for patient global assessment of disease activity, and 31.0 and 34.4 points for WOMAC function score. The PASS varied moderately across the tertiles of baseline scores but not across age, disease duration, or sex.
Conclusion: The use of PASS in clinical trials would provide more meaningful results expressed as a proportion of patients in an acceptable symptom state.

In clinical trials, at the group level, results are usually reported as mean and standard deviation of the change in score, which is not meaningful for most readers. The importance of incorporating patient perspectives in research into rheumatic diseases and defining outcomes that are comprehensive and influence clinical decision making was emphasised during the OMERACT 6 meeting.[1] Previous studies have dealt with the concept of the minimal clinically important difference (MCID)[2–4] or minimal clinically important improvement (MCII)[5] that could help in interpreting changes in scores in individual patients, by expressing the results as a proportion of improved patients. Another potentially clinically relevant concept is the patient acceptable symptom state (PASS), defined as the value beyond which patients consider themselves well. The MCID deals with the concept of improvement (feeling better) and the PASS the concept of wellbeing or remission of symptoms (feeling good). Thus, the PASS is undoubtedly a clinically relevant outcome for the patient.

The MCID and PASS concepts are complementary. If a patient with a high level of pain (90 mm on a 0–100 mm visual analogue scale (VAS)) experiences a decrease in pain of 40 mm, thus reaching 50 mm on the VAS, one can probably recognise a clinically relevant improvement (concept of MCID) but not a satisfactory state (concept of PASS). Results could be expressed both as a proportion of improved patients and of patients in a satisfactory state.

This prospective cohort study aimed at determining the PASS estimates for three main patient reported outcomes used in osteoarthritis (OA) trials[6]: pain, patient's global assessment of disease activity, and functional impairment.

MATERIALS AND METHODS

Study design

We conducted a 4 week prospective cohort study.

Study population

This study involved 1362 outpatients with knee and hip OA, as defined by the American College of Rheumatology,[7 8] included by 399 rheumatologists in France. Each rheumatologist had to include four patients, three with knee OA and one with hip OA. To be included in the study, patients had to experience pain from OA (≥30 mm on a VAS varying from 0 to 100), require treatment with a non-steroidal anti-inflammatory drug (NSAID), and be able to complete questionnaires in French. Inclusion could begin with the onset of treatment or a switch from one NSAID to another. Patients were excluded if they had a prosthesis on the assessed joint or if they had been treated by intra-articular injection in the 4 weeks before the study began. All patients initially visited the rheumatologist in charge of the patient, and an NSAID was prescribed (the drug and its dosage were chosen by the physician). A final visit to the same rheumatologist was scheduled 4 weeks later.

Table 1 shows the descriptive statistics on clinical and demographics variables.

Abbreviations: LDA, low disease activity; MCID, minimal clinically important difference; MCII, minimal clinically important improvement; NSAID, non-steroidal anti-inflammatory drug; OA, osteoarthritis; PASS, patient acceptable symptom state; VAS, visual analogue scale; WOMAC, Western Ontario McMaster Universities Osteoarthritis Index

Table 1 Baseline characteristics of patients

	Knee OA (n=914)		Hip OA (n=310)	
	Mean	SD	Mean	SD
Age (years)	67.8	10.2	65.7	10.8
Weight (kg)	75.2	14.2	72.2	14.0
Height (cm)	163.6	8.7	164.8	8.2
Body mass index (kg/m²)	28.1	4.7	26.5	4.1
Disease duration (years)	4.8	5.8	3.4	4.8
Pain score (0–100 mm VAS)				
Week 0	58.3	16.9	56.7	17.4
Change (week 0–week 4)	−24.5	22.1	−18.7	21.8
Patient global assessment (0–100 mm VAS)				
Week 0	58.7	19.1	58.6	16.5
Change (week 0–week 4)	−24.0	24.6	−19.5	23.5
WOMAC function score (0–100)				
Week 0	42.9	16.6	45.9	17.1
Change (week 0–week 4)	−11.6	14.4	−10.8	14.1
	No	(%)	No	(%)
Female sex	637	(69.7)	189	(61.0)
Kellgren and Lawrence grade				
II	178	(19.5)	57	(18.5)
III	394	(43.1)	145	(46.9)
IV	342	(37.4)	107	(34.6)
NSAID* intake during past 4 weeks	262	(28.8)	97	(31.3)
Analgesic intake†	513	(56.3)	209	(67.9)
Symptomatic slow acting drugs intake‡	311	(34.1)	123	(39.8)

*Non-steroidal anti-inflammatory drugs (before the start of the study); †other than NSAIDs (before the start of the study); ‡chondroitin sulphate, diacerhein, or avocado soybean unsaponifiables.

Measurements

The design of the trial included a baseline visit to the rheumatologist, a 4 week NSAID treatment phase, and a final visit at week 4. At the baseline visit, demographic and disease data (in particular, disease duration) were collected. Patients assessed their OA status at baseline and final visit. They assessed the following patient reported outcomes: (*a*) pain on movement during the 48 hours before the visit, measured on a 0–100 mm VAS; (*b*) global assessment of disease activity, measured on a 0–100 mm VAS; and (*c*) physical function, measured on the Western Ontario McMaster Universities Osteoarthritis Index (WOMAC) function subscale (17 items, five point Likert scale for each item; high scores indicate high degree of functional impairment; total score normalised to a 0–100 score).

At the final visit, patients' opinions of their state was also recorded by their answering "Yes" or "No" to "Taking into account all the activities you have during your daily life, your level of pain, and also your functional impairment, do you consider that your current state is satisfactory?".

Statistical analysis

All the analyses considered patients with knee and hip OA separately.

We used an anchoring method based on patient satisfaction with the current state. The same methods as for the MCII study (see companion paper in this issue[5]) were used, and the PASS was estimated by constructing a curve of cumulative percentages of patients as a function of the score of interest at the final visit among patients who considered their state satisfactory. Logistic regression was used to model the observations (fig 1). We targeted the point at the flattening of the curve at which most subjects stated they had a satisfactory status. This point corresponds to the 78.9th centile of the final score, and thus we propose to define the PASS as the 75th centile of the final score (at week 4), because it is very close to the point defined above and easier to derive. The model permitted us to determine that the target point was correctly approached by the 75th centile and to estimate the 95% confidence intervals. We also modelled

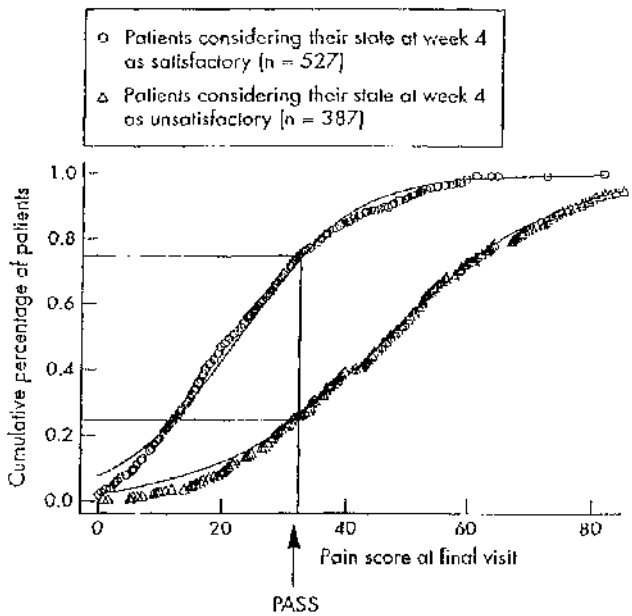


Figure 1 Aspects of the cumulative distribution function used to determine the PASS (pain scores in patients with knee OA). Among patients considering their state as satisfactory, 75% assessed their pain score at final visit below 32.3 mm on a 0–100 mm VAS (which is the PASS limit). Among patients considering their state as unsatisfactory, only 25% assessed their pain score at final visit below 32.3 on a 0–100 mm VAS.

the data from patients who considered their state unsatisfactory (fig 1).

In a second step, we stratified the analysis on the baseline score of interest (divided into tertiles) to assess whether the baseline scores for level of pain, patient's assessment of disease activity, and functional impairment affected the PASS estimates. That is we stratified (*a*) on the baseline pain score to estimate the PASS for pain; (*b*) on the baseline patient's assessment of disease activity score to estimate the

Table 2 Patient acceptable symptom state (PASS) scores according to location of OA

Patient reported outcomes	Knee OA		Hip OA	
	PASS	(95% CI)	PASS	(95% CI)
Pain (0–100 mm VAS), mm	32.3	(30.1 to 34.7)	35.0	(32.8 to 37.4)
Patient global assessment (0–100 mm VAS), mm	32.0	(29.5 to 34.8)	34.6	(32.3 to 37.1)
WOMAC function score (0–100)	31.0	(29.4 to 32.9)	34.4	(31.9 to 37.3)

PASS for patient's assessment of disease activity; and (c) on the baseline WOMAC function score to estimate the PASS for functional impairment.

In a third step, to investigate the effect of covariates (other than location of OA) on the PASS, we stratified the analysis successively by age, disease duration (both divided into tertiles), and sex.

Statistical analyses was performed with the SAS Release 8.2 statistical software package and the S plus 4.5 statistical software package.

Compliance with research ethics standards

This study was conducted in compliance with the protocol, good clinical practices, and the Declaration of Helsinki principles.

RESULTS

A total of 1362 patients were enrolled in the study: 1019 (75%) had knee and 343 (25%) hip OA; 913 (67%) were female; and the mean (SD) age was 67.2 (10.5) years. A total of 914 (90%) patients with knee and 310 (90%) with hip OA completed the final visit. Patients lost to follow up were excluded from the analysis and did not differ from completers in their baseline characteristics.

Among the completers, 527/914 (57.7%) patients with knee and 156/310 (50.2%) with hip OA considered their functional state at week 4 as satisfactory.

Table 2 lists the PASS estimates for the three patient reported outcomes and gives their 95% confidence intervals. For instance, patients with knee OA considered their state satisfactory if their pain score was less than 32.3 mm on the 0–100 mm VAS. The PASS estimates are similar (scores of approximately 33) across location of OA, whatever the patient reported outcome. These values are very close to those calculated for the 25th centile of the cumulative distribution function for the final score among patients with knee (fig 1) and hip (data not shown) OA who considered their functional state unsatisfactory.

Table 3 shows the estimates of the PASS stratified on the baseline score of interest. For instance, patients with knee OA with severe pain (high tertile of baseline pain score) considered their state satisfactory if their pain score was <27 mm on the 0–100 mm VAS. The PASS estimate varied moderately across tertiles of baseline scores (the higher the baseline score, the higher the PASS), but this trend is clearer for functional impairment.

The PASS estimates did not vary across age or disease duration tertiles or sex (data not shown).

DISCUSSION

In this prospective study we estimated the PASS for the three main patient reported outcomes used in clinical trials in knee and hip OA.

The PASS is the value beyond which patients consider themselves well. Thus, it can be considered a clinically relevant treatment target. It is an absolute value, not a change. Describing the number of patients achieving and maintaining such a state for a specified period of time will add useful information for daily practice and aid in the interpretation of trial and longitudinal results.[9]

This concept is very close to the low disease activity (LDA)[10 11] but applies only to patient reported outcomes (that is, symptoms). The LDA reflects an intermediate state between high disease activity and remission that could be called LDA or partial remission. An OMERACT 6 workshop focused on this concept for rheumatoid arthritis.[11] LDA was defined as a disease activity state deemed a useful treatment target by both physicians and patients. The definition of the PASS is anchored to the personal experience of the patient (satisfaction and adaptation to symptoms), although the LDA is anchored to both the patient's experience and the physician's clinical experience (treatment decision and prognosis). In a symptomatic disease such as OA, PASS and LDA are joined. In a disease such as rheumatoid arthritis, the PASS deals only with patient reported outcomes, although

Table 3 Patient acceptable symptom state (PASS) in patients with knee or hip OA, by baseline score of interest divided into tertiles

	Knee OA			Hip OA		
	Baseline score tertile			Baseline score tertile		
	Low	Intermediate	High	Low	Intermediate	High
Pain (0–100 mm VAS)	27.0 (24.6 to 29.9) (30 to 51.0)	34.5 (32.3 to 37.0) (51.1 to 66.2)	36.4 (33.2 to 40.0) (>66.2)	29.4 (26.0 to 33.7) (30 to 49.3)	35.2 (32.8 to 37.9) (49.4 to 65.4)	43.6 (38.9 to 49.6) (>65.4)
Patient's global assessment (0–100 mm VAS)	28.3 (25.5 to 31.6) (≤50.4)	34.3 (32.1 to 36.7) (50.5 to 68.7)	34.4 (32.1 to 36.7) (>68.7)	30.3 (26.6 to 34.9) (≤49.9)	33.5 (31.0 to 36.5) (50.0 to 69.9)	41.2 (37.2 to 45.9) (>69.9)
WOMAC function score (0–100)	20.4 (19.1 to 21.8) (≤35.3)	33.0 (31.3 to 34.9) (35.4 to 51.5)	43.1 (40.8 to 45.7) (>51.5)	20.6 (19.1 to 22.3) (≤38.7)	34.4 (31.9 to 37.4) (38.3 to 52.9)	44.2 (41.0 to 48.2) (>52.9)

Results are shown as the PASS (95% confidence interval) (tertile).

the LDA also encompasses factors such as biological signs of inflammation.

The concept of PASS is based on patient opinion as an external anchor, according to the OMERACT LDA module recommendation (the opinion based rather than data based approach seems more appropriate in deriving the LDA definition). The large sample of patients used as experts to determine remission in symptoms in our study is a good indicator of the representativeness.

The PASS was defined as the 75th centile of the final score in patients who considered their state satisfactory. This threshold relies on the data modelling and was chosen with the help of experts (NB, CB, DF, MH, DvdH, MD). However, these values are very close to those calculated for the 25th centile of the cumulative distribution function for the final score among patients who considered their functional state unsatisfactory. Thus, beyond the PASS limit were 75% of the patients who considered their current state satisfactory and only 25% of those who did not. Otherwise, the estimates of the PASS range from approximately 30 to 33 on a 0–100 point scale, whatever the patient reported outcome. The relevance of these results is reinforced by results which showed, in a study of patients who used intravenous patient controlled analgesia to self administer morphine sulfate after intra-abdominal surgery, that only 4% who rated their pain <30 mm on the 0–100 mm VAS requested additional analgesia, compared with 43–80% of those with pain scores of 31–70 or higher.[12] Thus, a pain score of <31 mm seems acceptable in this context as well.

In our study we investigated the effect of several covariates on the PASS estimates. The PASS varied moderately across the tertiles of baseline scores but less markedly than the MCII.[9] Thus, the PASS seems to be more robust than the MCII, which is affected by the initial level of symptoms, so the PASS is the recommended choice. However, the other factors investigated (age, sex, OA location, and disease duration) did not consistently modify the PASS estimates.

In conclusion, this study, dealing with a concept of emerging use, provides preliminary information facilitating the presentation and interpretation of results obtained in clinical trials. Further studies involving different datasets, clinical environments, languages, and countries, are necessary to validate these observations.

ACKNOWLEDGEMENTS

This study was supported by an unrestricted grant from Merck, Sharp & Dohme Chibret Laboratories, France.

Authors' affiliations

F Tubach, P Ravaud, G Baron, Institut National de la Santé et de la Recherche Médicale (INSERM) E 0357; Département d'Epidémiologie, Biostatistique et Recherche Clinique; Groupe Hospitalier Bichat-Claude Bernard (Assistance Publique—Hôpitaux de Paris); Faculté Xavier Bichat (Université Paris 7), Paris, France
B Falissard, Faculté de Médecine Paris-sud, Département de Santé Publique, Hôpital Paul Brousse (Assistance Publique—Hôpitaux de Paris), Villejuif, France
I Logeart, Merck, Sharp & Dohme Chibret Laboratories, Paris, France
N Bellamy, Department of Medicine, University of Queensland, Royal Brisbane Hospital, Brisbane, Queensland, Australia
C Bombardier, Institute for Work and Health, Toronto, Ontario, Canada
D Felson, Boston University School of Medicine, Boston, Massachusetts, USA
M Hochberg, University of Maryland, Baltimore, Maryland, USA
D van der Heijde, University Hospital, Maastricht, The Netherlands
M Dougados, Service de Rhumatologie B, Hôpital Cochin (Assistance Publique—Hôpitaux de Paris), Paris, France

REFERENCES

1 **Saag KG.** OMERACT 6 brings new perspectives to rheumatology measurement research. *J Rheumatol* 2003;30:639–41.
2 **Beaton DE, Bombardier C, Katz JN, Wright JG.** A taxonomy for responsiveness. *J Clin Epidemiol* 2001;54:1204–17.
3 **Juniper EF, Guyatt GH, Willan A, Griffith LE.** Determining a minimal important change in a disease-specific quality of life questionnaire. *J Clin Epidemiol* 1994;47:81–7.
4 **Jaeschke R, Singer J, Guyatt GH.** Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989;10:407–15.
5 **Tubach F, Ravaud P, Baron G, Falissard B, Logeart I, Bellamy N, *et al.*** Evaluation of clinically relevant changes in patient-reported outcomes in knee and hip osteoarthritis: the minimal clinical important improvement. *Ann Rheum Dis* 2005;64:29–33.
6 **Bellamy N, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P, *et al.*** Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. *J Rheumatol* 1997;24:799–802.
7 **Altman R, Alarcon G, Appelrouth D, Bloch D, Borenstein D, Brandt K, *et al.*** The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. *Arthritis Rheum* 1991;34:505–14.
8 **Altman R, Asch F, Bloch D, Bole G, Borenstein D, Brandt K, *et al.*** Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Association. *Arthritis Rheum* 1986;29:1039–49.
9 **Wells G, Anderson J, Boers M, Felson D, Haiberg T, Hewlett S, *et al.*** MCID/Low Disease Activity State Workshop: summary, recommendations, and research agenda. *J Rheumatol* 2003;30:1115–18.
10 **Boers M, Anderson JJ, Felson DT.** Deriving an operational definition of low disease activity state in rheumatoid arthritis. *J Rheumatol* 2003;30:1112–4.
11 **Wells G, Boers M, Shea B, Anderson J, Felson D, Johnson K, *et al.*** MCID/Low Disease Activity State Workshop: low disease activity state in rheumatoid arthritis. *J Rheumatol* 2003;30:1110–1.
12 **Bodian CA, Freedman G, Hossain S, Eisenkraft JB, Beilin Y.** The visual analog scale for pain: clinical significance in postoperative patients. *Anesthesiology* 2001;95:1356–61.

SECTION 9 - WOMAC Index: A Global Perspective

In mid-1982, the development of a valid, reliable and responsive international standard of measurement for OA clinical trials was just a concept. The difficulty of the challenge was unknown, but a high probability of failure was recognised. Developments in measurement, in other areas of rheumatology (particularly rheumatoid arthritis), the formation of OMERACT and OARSI, and the emergence of COX-2 class agents largely account for the unexpected acceleration of interest in outcome measurement in OA clinical trials. The precedent established by some of the early pioneers (John Lansbury, Otto Steinbrocker, Douglas Taylor), and more recent proponents (Robert Meenan, James Fries, Michel Lequesne) of outcome measurement in the rheumatic diseases, also propagated the discipline of musculoskeletal clinical metrology, and, in particular, raised awareness of methodological issues in health status questionnaire development. Nevertheless, in 1982, outcome measurement was far from being standardised, and the approach to outcome measurement in rheumatology clinical trials was highly variable, with respect to domains, variables, item specification and scales. Furthermore, there was no high-level international consensus on core set domains or preferred instruments, and no definition of what constituted a clinically important difference, or a symptom severity state acceptable to patients. The health status questionnaires of the time were generally available in English, or at best also in a few key European languages, and were not necessarily created using standard operating procedures, or culturally/linguistically appropriate methods.

The OA measurement environment, some 23 years later, in 2005, is completely different. It has been both interesting and a privilege to be an active participant in task forces, working groups and parties and consensus groups, that have shaped and standardised outcome measurement procedures for OA clinical research in rheumatology. These have included OMERACT, OARSI, WHO/ILAR, and IMMPACT.

The WOMAC Index is one of several assessment methods that have emerged. Its' importance in outcome measurement in diverse clinical research and clinical practice environments is best appreciated from reviewing the reference section of the latest version of the WOMAC User Guide (44). WOMAC User Guide VII is the latest in a series of WOMAC User Guides dating back to 1995. Different versions of the WOMAC User Guide have been distributed, since 1995, to over 10,000 WOMAC Index users. WOMAC User Guide VII contains references to over 1,000 publications (original papers, reviews, abstracts, and guidance documents) that draw on the WOMAC Index, and its applications, in various clinical research environments including the evaluation of anti-inflammatory drugs, COX-2 selective and specific agents, viscosupplements, orthopaedic surgical procedures, and physical forms of therapy. The Guide also details the more than 60 alternate-language translations of the WOMAC 3.1 series of questionnaires currently available, many available in different scaling formats to meet end-user needs and preferences.

The WOMAC Index has been produced in 5-point Likert, 11-point NRS and 100 mm VA formats, in electronic (e-WOMAC) and traditional paper-based applications, and with time frames that have varied as follows: 24-hours (original WOMAC 3.0), 48-hours (WOMAC 3.0 and WOMAC 3.1), last week (WOMAC 3.1W), last seven days

(WOMAC 3.1LSD), since last visit (WOMAC 3.1SLV), last month (WOMAC 3.0M). The WOMAC Index has been targeted on a single hip or knee joint (WOMAC 3.0 and WOMAC 3.1), on hips and/or knees (original WOMAC 3.0), on both knees together, and on both knees separately. In the case of injectable treatments, a version based on the knee to be injected has been developed (WOMAC 3.1IK). Finally the WOMAC has been produced in standard format as well as in signal (WOMAC 3.0S and WOMAC 3.1S), and short-form (SF-WOMAC 3.0) formats. These variations on the standard WOMAC Index have been produced to meet the specific needs of end users, and complement the standard 3.1 series WOMAC Index.

The innovative nature of the WOMAC Index was recognised by the Government of Canada in the formal registration of both the copyright to the source questionnaire and awarding the Canadian trademark (CDN No. TMA 545,986) to the author of this thesis (Appendix A).

Since 1996, the WOMAC 3.1 Index has been licensed for use in over 56 countries to major multinational pharmaceutical, biotechnology companies, academics, undergraduate and postgraduate students in various disciplines and to a large number of practitioners, particularly in the disciplines of rheumatology, orthopaedic surgery, physiotherapy and rehabilitation medicine.

The WOMAC Index is widely used in clinical research studies, and is now increasingly used in clinical practice environments. The Index often features as one of the most commonly used outcome measures in OA clinical studies reported at international rheumatology meetings. The WOMAC Index is identified as a measurement option by relevant organisations and regulatory agencies in Europe and the United States of America, including the draft guidelines of the European Medicines Agency (EMEA)(www.emea.eu.int), and the United States Food and Drug Administration (FDA)(www.fda.gov). The Index has been provided for use in Medical Research Council funded studies in the United Kingdom, Canada and Australia, and is one of the key clinical outcome measures for the National Institutes of Health Osteoarthritis Initiative (NIH-OAI) in the United States of America.

As noted in the preceding sections of this thesis, the WOMAC Index has played an important role in the specification of core set measures, and proposals for responder criteria and state-attainment criteria in OA(44). As previously noted, there are many different versions of the WOMAC Index, adapted for different applications, and which vary subtly in phraseology, language, time frame, joint target and platform.

The standard version, which is the WOMAC 3.1 series questionnaire, is joint targeted and uses a 48-hour time frame. It is this tri-dimensional, 24-item version, whether in 5-point Likert or 100 mm VA format, which has proven particularly popular and has met requirements in diverse clinical research and clinical practice requirements at a global level. The WOMAC Index development has permitted the detection of clinically important statistically significant differences in diverse research environments, and has arguably established itself as one of the international standards of measurement in OA clinical trials (44). The ability of the WOMAC Index to meet the measurement needs of a broad constituency of global users, by providing a valid, reliable and very responsive measure of patient reported outcome, meets the original measurement goals established in 1982. The influence of the WOMAC Index development, in supporting development of international measurement guidelines, core set clinical measures and proposals for

response criteria and state-attainment criteria, exceeds the original expectation, and has been achieved through international collaboration with colleagues in academia, industry, regulatory bodies and clinical practice. The success of the WOMAC Index is, in no small part, attributable to the original 100 patients with OA, who were interviewed in 1983 and 1984, and whose careful explanation of the impact of OA on different aspects of their everyday life, shaped the content and format of the WOMAC Index. Without their insight and perseverance, it would not have been possible to develop a valid measure meeting a global need.

Reference:

44. Bellamy, N. WOMAC Osteoarthritis Index – User Guide VII (Published by N. Bellamy, Brisbane, Australia) 2005.

WOMAC OSTEOARTHRITIS INDEX

USER GUIDE VII

Nicholas Bellamy

CONTENTS

FOREWORD

The WOMAC Osteoarthritis Index is a tridimensional, self-administered, patient-centered health status questionnaire. Its item inventory has been designed to capture the essential elements of pain, stiffness and physical disability in patients with osteoarthritis of the knee and/or hip joints. The WOMAC Index has been subject to more than 20 studies examining its basic clinimetric properties and has been translated into over 60 different language forms. Its ability to detect change in health status has been demonstrated following patient exposure to a variety of different interventions: Nonsteroidal Anti-inflammatory Drugs, COX-2 Inhibitors, Analgesics, Viscosupplements, Physiotherapy, Orthopaedic Surgery.

WOMAC User Guide VII provides updated information on Index development and validation, and includes information about recently available alternate-language translations, electronic data capture using the WOMAC Index, and recent experience with the Index in clinical research and clinical practice environments. This release of the WOMAC User Guide also contains an explanation of the meaning of questions in the WOMAC Index inventory. WOMAC User Guide VII provides reflections on recent developments in the area of response status assignment in OA based on change criteria (OARSI, OMERACT-OARSI, MPCI, MCII and WOMAC 20 50 70 responder criteria), and state attainment criteria (MCAS, PASS, and BLISS Index), developed in hip and/or knee OA environments. Finally WOMAC User Guide VII contains summaries of two other related patient-centered health status measures, the Australian/Canadian (AUSCAN) Osteoarthritis Hand Index, and the Osteoarthritis Global Index (OGI).

I gratefully acknowledge the generous collaboration of the many research associates, technical experts, rheumatologists, orthopaedic surgeons, epidemiologists and biostatisticians who facilitated the successful development of the WOMAC Index. I am particularly grateful to my colleague, Jane Campbell, who has worked diligently over several years on WOMAC projects.

Nicholas Bellamy MD MSc MBA FRCP(Glas) FRCP(Edin) FRCP(C) FACP FAFRM FRACP
Professor of Rehabilitation Medicine
Director - Centre of National Research on Disability and Rehabilitation Medicine (CONROD)
The University of Queensland
Faculty of Health Sciences
Mayne Medical School
Herston Road, Herston
Queensland 4006
AUSTRALIA

Tel +61 7 3365 5558
Fax +61 7 3346 4603
e-mail nbellamy@medicine.uq.edu.au
WOMAC Website: www.womac.org

WOMAC USER GUIDE VII

OVERVIEW

The WOMAC Osteoarthritis Index is a tri-dimensional, disease-specific, self-administered, health status measure (1-64). It probes clinically-important, patient-relevant symptoms in the areas of pain, stiffness and physical function in patients with osteoarthritis of the hip and/or knee. The index consists of 24 questions (5 pain, 2 stiffness, 17 physical function) and can be completed in less than 5 minutes. It is available in Likert (WOMAC LK3-series), Visual Analogue (WOMAC VA3-series) and Numerical Rating (WOMAC NRS-series) scaled formats. WOMAC is valid, reliable, and sufficiently sensitive to detect clinically-important changes in health status following a variety of interventions (pharmacologic, surgical, physiotherapy, etc). It has been translated into many different languages and has been requested for use by more than 500 researchers in over 50 different countries (65-1041). Several different formats of the WOMAC Index have been produced including the WOMAC Original, 3.0, 3.0S, 3.1, 3.1S, 3.1W, 3.1M and 3.1SLV, 3.1(IK) SF-WOMAC and e-WOMAC. The WOMAC Index has become a standard measure for clinical trials in hip and knee osteoarthritis. Several major agencies and organisations, as well as individual researchers, have given focus to the way in which measurement should be conducted and standardised techniques and tools from which to choose (1042-1073).

CONCEPTUAL BASIS

The conceptualisation of WOMAC began in 1981, in my MSc thesis (29), and at a time when there were no international standards of clinical measurement in osteoarthritis (OA) clinical trials (4). The need for such an instrument arose from a review of the clinical trials literature (4). In that review of 63 NSAID trials in OA, pain had been assessed in 58, patient global assessment in 51, range of movement in 45, physician global assessment in 42, joint stiffness in 35, qualitative aspects of sleep in 28, walk time in 23, activities of daily living in 22, joint tenderness in 19, analgesic consumption in 15, joint swelling in 15, signal joints in 10, ascent time in 3, muscle power in 3, hand function in 3, radiology in 2, and joint temperature in 1. Not only did the variables differ, but there was considerable variability in the scales and instruments employed. Apart from physician and patient global assessments, pain and stiffness were the only other two variables monitored in more than half of the trials. While the second most important symptom of OA, physical disability, was monitored in only 35% of the studies, measures of physical function detected a significant difference between an active treatment phase and either a washout phase or placebo treatment period in 86% of assessable studies. Most of the other measures used in the 63 NSAID trials were observer-dependent, i.e., required a physician or allied health professional to make a judgement based on observed performance or clinical examination. Given that such measures are subject to observer variability and may lack clinical importance to individual patients, we elected to focus on three observer-independent, patient-relevant measures, i.e., pain, stiffness, and physical function. An index applicable to OA hip and knee patients was considered necessary given the propensity for OA to affect these two anatomic areas, and the fact that most

pharmacologic studies have focused on these two joints.

DERIVATION OF THE ITEM INVENTORY

In order to construct the item inventory of WOMAC, the dimensionality of the symptomatology of OA was explored in 100 patients with hip and/or knee involvement (4). The survey questionnaire was developed by a peer review process involving the opinions of four academic rheumatologists and two clinical epidemiologists experienced in clinical measurement in the rheumatic diseases. Initial questions were open-ended and probed the clinical importance and characteristics of any pain, stiffness, physical, social or emotional dysfunction. Once spontaneous responses to these questions were exhausted, a battery of closed-ended questions, derived from (and modified where necessary) six existing questionnaires (Health Assessment Questionnaire - HAQ, Functional Status Index - FSI, Arthritis Impact Measurement Scales - AIMS, Pooled Index, McMaster/Toronto Assessment Index - MACTAR, McMaster Health Index Questionnaire - MHIQ), was used to complete the assessment of each dimension and quantitate any sources of discomfort and disability. The survey questionnaire was administered by face to face interview. The following data were recorded:

1) The presence or absence of each of several types of discomfort or disability.

2) The frequency with which each type of discomfort or disability occurred (daily, weekly, fortnightly, monthly or less).

3) The importance of each type of discomfort or disability to the patient (0 = none, 1 = slight, 2 = moderate, 3 = very, 4 = extreme).

Patients were specifically questioned about sources of discomfort and disability recently experienced and attributed to OA in the hips and/or knees. They were questioned regarding the perceived importance of each type of discomfort and disability in order to assess its clinical relevance. Gender-specific questions relating to physical disability (e.g. ironing) were avoided and questions phrased in more general terms (e.g. light domestic duties). Questions relating to sexual disability were not included since this has been previously noted to inhibit responses to subsequent non-sexual questions. Although questions relating to social and emotional dysfunction were included in the evaluation of the dimensionality of the symptomatology of OA, and were subsequently included in two validation studies of the WOMAC Index, they do not form part of the WOMAC 3.1 Index and will not be considered further. The social dimension was deleted because of poor construct validity. In contrast, the emotional dimension performed extremely well but was not included, because at that time it seemed redundant without the accompanying social dimension. However, the recent Initiative on Methods, Measurement and Pain Assessment in Clinical Trials (IMMPACT) consensus included not only recommendation to measure pain and function, but also the inclusion of measurement of emotional function (1058,1072) As a consequence, the WOMAC emotional function subscale may be reintroduced for clinical trials and clinical practice applications in hip and knee osteoarthritis which require measurement of the emotional domain.

The prevalence (P), mean importance score (MIS), and percentage of symptomatic patients experiencing daily or weekly symptoms (DW) varied between component

items. The range of values for items retained within the WOMAC Index was as follows: Pain: P = 56-77%, MIS = 2.51 - 2.63, DW = 94 - 96%; Stiffness: P = 47 - 73%, MIS = 2.30, 2.52, DW = 100%; Physical function: P = 33 - 70%, MIS = 2.26 - 2.67, DW = 71 - 100%. No association was observed between MIS for pain, stiffness or physical function and the following variables: age, gender, disease duration. The item inventory of the WOMAC Index was directly derived from the aforementioned study. This development strategy ensured that the Index had face and content validity and that it probed symptoms which occurred commonly, were regarded as being of importance to symptomatic individuals, and generally were experienced on a daily or weekly basis. The fact that the symptoms occurred daily or weekly was important, since patients can only record change on an index if they have had opportunity to re-experience events at a time when the effects of a new intervention might be expected to have occurred. Thus the item inventory of the WOMAC Index monitors clinically-important events, relevant to patients and recurring with high frequency in symptomatic individuals.

VALIDATION STUDIES (Reliability, Validity, Responsiveness)

Two major validation studies of the WOMAC Index have been completed (14,15). The goal was to assess the reliability and validity of the Index, and evaluate its responsiveness to two different forms of intervention (orthopaedic and pharmacologic). For the purpose of validation, responses to WOMAC questions were scaled in two different formats. The LK-scaled version allowed patients to make their responses on 5-point adjectival scales (0 = none, 1 = mild, 2 = moderate, 3 = severe, 4 = extreme). In contrast, the VA-scaled version permitted responses to be made on 100 mm horizontal visual analogue scales with end markers, outwith which were placed the following descriptive anchors (left end = none, right end = extreme).

Orthopaedic Validation Study

This study employed a quasi-experimental, one-group repeated measures design (14) in 30 OA patients undergoing total arthroplasty of the hip (n=16) or knee (n=14). Patients were evaluated the day before surgery and at 6 weeks, 3 months and 6 months post-operatively. In addition to LK- and VA-scaled versions of WOMAC, the following measures were administered concurrently for validation purposes: 1) Modified Doyle Index, 2) Lequesne Index, 3) Bradburn Index of Well Being, and 4) Social Component of the MHIQ. Additional measures included interviewer global assessment, patient global assessment, 50' walk time, joint range of motion, intermalleolar straddle, and intercondylar distance.

Validity testing was based on construct validity determined using Pearson's correlation coefficients. Correlation coefficients, based on the WOMAC, Doyle, Lequesne, Bradburn, and MHIQ Indices, supported the construct that the pain, stiffness and physical function subscales were valid. In particular, WOMAC Index items showed statistically significant correlations with other indices probing the same dimension (i.e., convergent construct validity). Furthermore, WOMAC Index items showed higher levels of correlation with indices probing the same dimension than with indices probing other dimensions (i.e., divergent construct validity). Construct

validity was demonstrated for both the LK- and VA-scaled versions of the index.

Reliability was determined by Cronbach's alpha pre-operatively, and at 6 weeks and 6 months post-operatively. Reliability values for the three subscales at the aforementioned times were as follows: Pain: LK = 0.80, 0.78, 0.93; VA = 0.88, 0.88, 0.93; Stiffness: LK = 0.88, 0.75, 0.88; VA = 0.87, 0.73, 0.96; Physical Function: LK = 0.93, 0.92, 0.97; VA = 0.88, 0.91, 0.94. These data suggest that both the LK- and VA-scaled versions are of excellent reliability.

Responsiveness was evaluated by examining the p values generated using Wilcoxon's test comparing 6 month post-operative versus pre-operative status. Despite the small sample size (n=30), statistically significant, clinically-important improvements in health status were noted on all three subscales. Statistical p values, based on summated subscale scores at 6 months post-operatively for the LK- and VA-scaled versions, respectively, were as follows: Pain: LK = ≤ 0.001; VA = ≤0.001; Stiffness: LK = <0.001; VA = <0.001; Physical Function: LK =≤0.001; VA = <0.001. These data attest to the responsiveness of the Index even when monitoring effects in small groups of patients. It was noted in this study that physical function item 18 (getting on/off a bus) was less responsive than other items, largely due to the fact that very few individuals ever travelled by bus. It was not included in the pharmacologic validation study or the final English-language version of the original Index. However, in some parts of the world, bus travel is common, and for this reason its use has been retained, as an alternate travel item, in some later forms of the Index.

Pharmacologic Validation Study

The study employed a double-blind, randomized, controlled trial design (15) and compared two NSAIDs [Isoxicam (n= 28), Piroxicam (n = 29)] in 57 patients with OA hip (n=18) or knee (n=39). Patients were evaluated at enrolment and again one week later without any change in therapy in order to obtain test-retest reliability estimates at steady state. Thereafter, patients underwent a one-week NSAID-free washout period. Finally they were evaluated after 2, 4 and 6 weeks of active treatment. In addition to LK- and VA-scaled versions of WOMAC, the following measures were administered concurrently for validation purposes: 1) Modified Doyle Index, 2) Lequesne Index, 3) Bradburn Index of Well Being, and 4) Social Component of the MHIQ. Additional measures included interviewer global assessment, patient global assessment, 50' walk time, total range of motion, and intermalleolar straddle. One of the major strengths of the validation design was that randomization created two groups similar in baseline characteristics and response potential. Independent evaluations of reliability, validity and responsiveness were undertaken on these two separate treatment groups.

Test-retest reliability by Kendall's tau c statistic using a one-week retest interval was as follows: Pain (combined groups): LK = 0.68, VA = 0.64; Stiffness (combined groups): LK = 0.48, VA = 0.61; Physical Function (combined groups): LK = 0.68, VA = 0.72. Internal consistency by Cronbach's alpha was as follows: Pain: LK - Isoxicam = 0.86, Piroxicam = 0.89; VA - Isoxicam = 0.81, Piroxicam = 0.73; Stiffness: LK - Isoxicam = 0.90, Piroxicam = 0.91; VA (not determined); Physical Function: LK - Isoxicam = 0.95,

Piroxicam = 0.95; VA - Isoxicam = 0.91, Piroxicam = 0.89. These data suggest that both the LK- and VA-scaled versions are of excellent reliability. Test-retest coefficients are in the mid range due to interval change occurring in health status of participating patients over the one-week interval.

Validity testing was based on construct validity determined using Pearson's correlation coefficients. Correlation coefficients, based on the WOMAC, Doyle, Lequesne, Bradburn, and MHIQ Indices, supported the construct that the pain, stiffness and physical function subscales were valid. In particular, WOMAC Index items showed statistically significant correlation with other indices probing the same dimension (i.e., convergent construct validity). Furthermore, WOMAC Index items showed higher levels of correlation with indices probing the same dimension than with indices probing other dimensions (i.e., divergent construct validity). Construct validity was demonstrated for both the LK- and VA-scaled versions of the Index.

Responsiveness was evaluated by examining p values generated using Wilcoxon's test comparing end of washout values with end of 6-week treatment values. Despite the small sample sizes (Isoxicam = 28, Piroxicam = 29), statistically significant improvements in health status were noted on all three subscales. Statistical p values, based on aggregated subscale scores after 6 weeks of active treatment for the LK- and VA-scaled versions, respectively, were as follows: Pain: LK - Isoxicam = ≤0.001, Piroxicam = ≤0.003, VA - Isoxicam = ≤0.001, Piroxicam = ≤0.003; Stiffness: LK - Isoxicam = ≤0.001, Piroxicam = ≤0.013, VA - Isoxicam = ≤0.001, Piroxicam = ≤0.013; Physical Function: LK - Isoxicam = <0.003, Piroxicam = ≤0.002, VA - Isoxicam = <0.003, Piroxicam = ≤0.002. These data attest to the responsiveness of the WOMAC index in two separate but similar small groups of OA patients.

On the basis of these two validation studies, the final version of WOMAC was established. Since LK- and VA-scaled formats of the Index had been separately validated, two versions of the Index were produced: WOMAC LK 3.0 and WOMAC VA 3.0. These versions are identical with respect to item inventory and differ only in the scales on which patients respond to the component questions.

SPECIFIC CLINIMETRIC ISSUES

The aforementioned validation studies established the reliability, validity and responsiveness of the Index. Face and content validity were ensured by the method used to develop the item inventory, while construct validity was demonstrated against the Doyle, Lequesne, Bradburn and MHIQ indices. Having established the basic clinimetric properties of the Index, we wished to address a number of specific issues in order to more fully understand the robustness of the instrument (27).

Likert vs Visual Analogue Scaling

In both major validation studies, (14,15) WOMAC VA 3.0 was slightly more sensitive than WOMAC LK 3.0, as judged by the size of the p values generated in parallel analyses of concurrent responses made by the same patients. It is of note that in a subsequent post validation reapplication of WOMAC VA 3.0 we were able to detect significant differences between two NSAIDs on the pain and stiffness subscales, no difference being noted in physical function (21). Although there has been much debate in the literature as to which of these two scales

is to be preferred, high levels of correlation have generally been observed between scores made concurrently on LK and VA scales. Although some investigators have reported difficulty amongst patients using VA scales in general, to date we have not encountered any difficulty with patients comprehending or completing WOMAC VA 3.0.

Since the development of the WOMAC Index, we have received requests for an NRS scaled version. We have created an NRS version from existing LK and VA scaled versions. Although we have not formerly validated the NRS scaling format, previous experience with NRS scales in OA (1016, 1074) and RA (1075), suggests that the NRS version may offer the simplicity of LK scaling with similar responsiveness to VA scaling. As a consequence both WOMAC and AUSCAN Indices (1076) Indices have been made available in LK, VA and NRS scaling formats.

Prior Score Availability

Whether patients should be shown their prior scores when repeatedly self-assessing health status remains controversial. We have had experience using the Index under both blind and informed conditions. This issue was addressed more formally as part of the pharmacologic validation study (n=28) (15). At the end of the study, patients completed WOMAC blind (i.e., on a blank questionnaire without reference to any previous data). Then, after having completed several other questionnaires, they were given a second opportunity to rate their current WOMAC status on the same WOMAC questionnaire on which they had previously marked their end of washout scores at the beginning of the study. A comparative analysis showed no clinically important or statistically significant differences between the results obtained from the two different forms of administration (20). This same issue was addressed using WOMAC VA3.0 in a double-blind, randomized, controlled trial comparing flurbiprofen SR and diclofenac SR in 70 patients with OA knee (25). No statistically significant differences were noted between the severity scores at termination under blind versus informed administration. The item scores for the two forms of administration were highly correlated. We, therefore, feel that the Index can be administered either with or without prior score availability. It is traditional at the present time to generally administer such questionnaires without prior score availability.

Time Frame Dependency

Many questionnaires fail to specify the time frame over which patients should consider the severity of their symptoms. Others specify time frames which exceed the interval between reassessments in the study. There are few published data on whether changing the time frame alters data interpretation. It has been demonstrated, however, that if the time frame is very short (e.g. "now"), circadian variation may affect the results obtained. In contrast, if the interval is too long, then recall (i.e., pain memory) may be faulty. We have performed an evaluation of the time frame dependency of WOMAC VA 3.0 in a small group (n=19) of patients entered into a randomized, controlled trial of two NSAIDs (40). At the final assessment, patients were asked to complete in random order three versions of WOMAC differing only in the time frame over which they were asked to consider their symptoms. The three time frames were: previous 24 hours, previous 48 hours, and previous 2 weeks. Although limited by the

relatively small number of patients, we observed no time frame dependency of questionnaire responses over the 14-day period. We feel justified in varying the time frame over which questions are asked (at least between 1-14 days) when using the WOMAC Index, depending on the dynamic requirements of any given study. We have most frequently selected an interval of the previous 48 hours.

Signal vs Aggregate Measurement

There are a number of reports in the literature suggesting that signal methods of measurement may be realistic alternatives to more comprehensive methods. A signal is a symptom or sign of disease which acts as the sole focus of measurement. Such signals may be individualized (i.e., tailored to the symptom profile of individual patients). By appropriate signal selection, it may be possible to improve the efficiency of the measurement process by restricting measurement to aspects of disease that are clinically important and have good response potential. We have conducted two such studies, one based on WOMAC LK 3.0 (n=30) (13), and the other on WOMAC VA 3.0 (n=70) (1). We call these alternate versions of the Index WOMAC LK 3.0S and WOMAC VA 3.0S, respectively. They are identical to the parent Index with the exception that on the last page, having completed the Index, patients are instructed as follows: "**Now** we would like you to think again about each of the aforementioned symptoms which you have just rated. Then select **one** pain item, **one** stiffness item, and **one** physical function item which are most important to you, i.e., which you most hope the treatment you are about to receive will **improve**. Indicate your selections by circling the appropriate item numbers. Remember to select only **one** pain item, **and one** stiffness item, **and one** physical function item".

In both studies a signal analysis, based on one pain, one stiffness and one physical function item per patient, was compared against the traditional, or aggregate analysis, in which three separate subscale scores were created by summation of component item scores (i.e., 5 pain, 2 stiffness, 17 physical function). The signal method was capable of detecting statistically significant improvements in both studies. As expected, patients varied considerably in which items they selected, not all signals selected being the most severely affected items. In both studies the statistical efficiency of the signal approach was slightly greater than that of the aggregate approach and was attended by smaller sample size requirements. However, there are two limitations to the signal approach. Firstly, in both studies we noted (albeit at a low prevalence) the occurrence of deterioration in non-signal items. Secondly, in our study with WOMAC VA 3.0S, we administered the Index at the beginning and end of the trial, and noted that 80% of patients had switched signal selection by the end of the trial (1). It was unclear from the analysis why this occurred, but the data suggested that patients may have selected new signals on the basis of a lack of improvement in their initial signal selections during the course of the trial such that the new signals had become subjectively more important. We are concerned that signal selection may not be stable. At the present time the signal approach to measurement remains experimental and we only recommend use of the complete Index, i.e., WOMAC LK 3.1, WOMAC VA 3.1 and WOMAC NRS 3.1 .

Parametric vs Non-Parametric Analysis

There is some controversy as to whether ordinal level data can be analyzed using parametric techniques. There is also debate as to the nature of the VA scale and whether it should be analyzed by non-parametric techniques. Parametric techniques, for comparisons of continuous data, generally require a normal distribution. To address some of these issues, we have compared the results of parametric (Student's t-test) and non-parametric (Wilcoxon's test) analysis of data from the WOMAC validation studies (14,15) In both studies we observed that in many instances parametric and non-parametric treatments of the data were in agreement. Overall, non-parametric methods provide a more conservative estimate of the response. Therefore, where data are normally distributed, parametric methods may be considered appropriate. However, where they are not normally distributed, non-parametric methods should be used.

Relative Efficiency

The relative efficiency (RE) statistic is one estimate of the comparative responsiveness of different instruments. It is defined by the square of the ratio of two t values (or z values), e.g. RE (WOMAC vs HAQ) = (t_{WOMAC}/t_{HAQ}). If the RE is >1.0, then the instrument in the numerator can be inferred to be the more responsive measure of outcome, requiring smaller sample sizes and/or detecting smaller effect sizes than the instrument in the denominator. The RE value is rarely unity and yet there are no standards defining the significance of values <1.0 or >1.0.

We have examined the RE of WOMAC versus other measures in four separate studies. The first and second studies were the validation studies previously reported (14,15). The third was a double-blind, randomized, controlled, clinical trial of meclofenamate versus diclofenac sodium in OA knee (21), and the fourth a quasi-experimental, one-group repeated measures trial of total knee arthroplasty (26). While we have no data on the RE of the stiffness subscale against other stiffness measures, our data on the relative efficiency of the WOMAC pain and physical function subscales are as follows: WOMAC pain vs HAQ pain = 1.59, AIMS pain = 0.81, Lequesne pain = 0.81, Doyle Index = 1.09. WOMAC physical function vs HAQ physical function = 1.13, AIMS physical function = 1.75, Lequesne maximum distance walked = 1.65, Lequesne activities of daily living = 1.14, Walk time = 1.2, 1.3, 3.7, 3.4, Intermalleolar straddle = 4.8, 5.1, 4.9, 2.3, Intercondylar distance = 6.9, 3.3, range of motion = 1.3, 1.4, (132.0), (67.9). These data suggest that the WOMAC Index is more efficient than measures of walk time, intermalleolar straddle, intercondylar distance, and range of motion. Furthermore, RE values for WOMAC are >1.0 for 91% of the aforementioned comparisons and suggest that overall the WOMAC may be a more efficient index for assessing outcomes in OA clinical trials.

In a study by Theiler et al a German version of the WOMAC appeared to be more responsive than the Lequesne Index of Clinical Severity (1021).

Recent studies, comparing the WOMAC against several different generic health-related quality of life (HRQOL) measures, have generally concluded that the WOMAC Index is more efficient in detecting disease-specific changes than the HRQOL measure (66,80,92,103). The combination of WOMAC and a generic HRQOL index would

be ideal, in those situations where the measurement goal is to evaluate the impact of an intervention on both the patient's disease (WOMAC) and the patient as a whole (HRQOL), since these two different approaches to health status assessment are mutually complementary.

Weighting and Aggregation

While the three WOMAC subscales are usually analysed separately, there are occasions when it is desirable to aggregate them together into a single score. To understand the options available for weighting and aggregation, it is necessary to consider some basic properties of the Index. In particular, it should be noted that all scales have a base score of zero. The scale lengths of the different subscales and versions of the index differ as follows: WOMAC LK3.1- Pain = 20, Stiffness = 8, Physical Function = 68; VA3.1- Pain = 500, Stiffness = 200, Physical Function = 1700, NRS3.1- Pain = 50, Stiffness = 20, Physical Function = 170. In general, the pain subscale is more responsive than the physical function subscale, and the stiffness subscale assumes an intermediate position. In order to assess some of the implications of weighting and aggregation, we have examined the importance of individual symptoms, the relative importance of pain vs stiffness vs physical function, and the relationship between different symptoms in each of the subscales.

Our original evaluation of the dimensionality of the symptomatology of OA suggested that symptomatic patients tended, on average, to regard their symptoms as moderately important (3) We subsequently studied a small group of OA knee patients (n=17) participating in a double-blind, randomized, controlled, clinical trial comparing flurbiprofen SR and diclofenac sodium SR using WOMAC VA3.0 (25) At the end of the 6-week trial, participants completed WOMAC VA 3.0 and shortly thereafter completed an alternative form of WOMAC on which they were asked to separately rate on 10 cm VA scales the importance which they attached to being completely symptom free on each of the 24 component items. No statistically significant correlation was noted between severity and importance scores. Except for two items, no significant difference was noted between the scores of individual items and the average score for the subscale to which that item belonged. The level of interitem correlation for components on each of the three subscales was high: pain = 0.79-0.96, stiffness = 0.83, physical function = 0.52-0.98. Principal component analysis showed that Factor I accounted for 88% of the variance in pain and 83% of the variance in physical function. The factor loading was high on each individual pain item (0.92-0.95) and each individual physical function item (0.70-0.97). There was relatively little additional variance accounted for by Factor II (pain =7%, physical function = 6%). Thus, analysis of individual WOMAC items within a subscale suggested that, although highly correlated, they measured different aspects of the disease. Factor analysis supported the contention that scores from items within a subscale could be summated into subscale scores.

We have evaluated the use of a device called the "Patient Assessment of the Relative Importance of Symptoms (PARIS) sectogram" (26). The sectogram is a mechanical device consisting of three interlocking, laminated 360° discs (red = pain, yellow = stiffness, blue = physical function), which are riveted together. Patients are presented the disc, set to display equal 120° segments for pain, stiffness, and physical function. The patient is

then instructed to rate the relative importance of being free of pain versus free of stiffness versus free of physical disability by resetting the relative sizes of the three sectors. The patient's preferences are scored by reading off the number of degrees displayed on the periphery of the sectogram. The mean relative importance values at baseline, expressed as percentages, in 54 patients studied in a 12-week, double-blind, randomized, controlled, clinical trial of tenoxicam vs diclofenac (12), were as follows: pain = 42%, stiffness = 21%, physical function = 37%. Reliability coefficients, based on test-retest at baseline and termination in the trial, for this procedure were high: pain = 0.86, 0.90; stiffness = 0.80, 0.88; and physical function = 0.81, 0.83.

Given the aforementioned properties of the index, there are five alternative approaches to weighting and aggregation.

A) The simplest and most commonly employed approach is by simple summation of the 24 component item scores (Total WOMAC Score). This provides a single value weighted according to differential scale lengths (ratio = 5:2:17) but without any correction for the relative importance of the different subscales.

B) Alternatively, a normalization procedure can be used to correct for differences in scale length (Normalized WOMAC Score). This is similar to the procedure employed in a normalization of the Arthritis Impact Measurement Scales (AIMS).

In order to normalize the LK 3.0 Index on 0-10 scales, the following correction factors are used where S = sum of raw scores of items in dimension: Pain -Score Range: 0-20, Normalization = (Sx0.50); Stiffness - Score Range: 0-8, Normalization = (Sx1.25); and Physical Function - Score Range: 0-68, Normalization = (Sx0.147).

In order to normalize the VA 3.0 Index on 0-100 scales, the following correction factors are used where S = sum of raw scores of items in dimension: Pain -Score Range: 0-500, Normalization = (Sx0.20); Stiffness - Score Range: 0-200, Normalization = (Sx0.50); and Physical Function - Score Range: 0-1700, Normalization = (Sx0.059).

In order to normalize the NRS 3.0 Index on 0-10 scales, the following correction factors are used where S = sum of raw scores of items in dimension: Pain -Score Range: 0-50, Normalization = (Sx0.20); Stiffness - Score Range: 0-20, Normalization = (Sx0.50); and Physical Function - Score Range: 0-170, Normalization = (Sx0.059).

Once subscale values have been normalized, they can be summated to provide a single value in which the three component subscales are equally weighted (i.e., ratio 1:1:1). This procedure makes no correction for the relative clinical importance of different subscales. Since all items are individually scored on the same scales, averaging scores has a similar effect to normalization unless the normalized scale length differs from the original.

C) Another method of bringing scores on the three subscales to a uniform scale length is that employed in the construction of the Pooled Index. The correction is made by first calculating derived units (DU) for scores on each of the three subscales. For example, DU_{PAIN}=Change score for pain ÷ Standard deviation of the change score. The Pooled WOMAC Score = DU_{PAIN} + $DU_{STIFFNESS}$ + $DU_{PHYSICAL\ FUNCTION}$. This procedure does correct for differences in scale length but not

for the relative clinical importance of the different subscales.

D) As noted previously, we have recently started to address the relative clinical importance of different types of symptoms to individual patients. While this work is still in progress, our preliminary observations suggest that individual patients vary in the relative importance they assign to being free of pain vs free of stiffness vs free of physical disability. The PARIS sectogram can be used in future studies to derive individualized weights which could be factored into the data analysis. Alternatively, it is possible to use our aforementioned weights derived from a group of 54 OA patients. The weighting factors are as follows: Pain = 0.42, Stiffness = 0.21, Physical Function = 0.37. While these weights can be applied at a subscale level in deriving an Importance- Weighted Total WOMAC Score, they might be more appropriately applied in deriving Importance-Weighted Normalized WOMAC or Pooled WOMAC Scores. We are continuing to investigate the advantages and restrictions of this approach. Our current data are limited and, therefore, a firm recommendation cannot be made at the present time. Nevertheless, weighting, according to the relative importance of symptoms, offers some interesting opportunities.

E) A final approach to combining information is in a set of response criteria. In this approach the success versus failure of treatment in individual patients could be adjudicated by the occurrence (or non-occurrence) of changes exceeding a certain magnitude on a defined combination of measures. We are currently interested in the possibility of using the WOMAC Index as the core of a set of response criteria for OA clinical trials.

We have examined the relative efficiency (RE) of the Total WOMAC Score and the Pooled WOMAC Score and observed that the weighting and aggregation procedures may tend, overall, to maintain or possibly increase relative efficiency (14,15,21).

Computerization

We have previously validated a computerized version of WOMAC VA 3.0 against the original paper version (18). Thirty patients with OA knee completed both forms of the Index. Patients were instructed to indicate their scores directly on the computer screen for each of the 24 WOMAC items using a "mouse" to move the cursor. An opportunity was provided to review cursor placements and modify where necessary. The corresponding numerical values between 0 and 100 were automatically generated for each VA scale from the exact cursor placements, but were not revealed to the patient. The data were downloaded each night to a central computer 200 km away. Correlation analysis demonstrated excellent criterion validity against the original paper index: pain = 0.90, stiffness = 0.87, physical function = 0.97. Despite wide variation in computer literacy skills, all patients completed the task and the data were successfully transmitted to the central computer. These data suggest that the computerized version of WOMAC VA3.0 is a valid alternative to the paper version. The results of this study provide exciting opportunities for future trials, particularly multicentre studies using remote data entry terminals. Since this study, the software has been redesigned to run on PC computers.

We have recently evaluated Likert and NRS touch screen versions of the WOMAC Index. Initial results have been favourable (47-49).

Alternate Language Translations

Originally developed in the English language, the WOMAC Osteoarthritis Index has been translated into many different languages. A list of the more than 60 authorised alternate language translations, currently available through our office, is provided in Appendix I. Additional authorised alternate language translations of the WOMAC VA3-series WOMAC LK 3-series and WOMAC NRS 3-series are being planned at the present time. **It is strongly recommended that only original authorised alternate-language translations of the WOMAC Index be used in clinical research and clinical practice applications.**

Rasch Analyses

A number of Rasch analyses of WOMAC data have been performed (33,45,61,62). In a Rasch analysis of 655 OA patients Wolfe and Kong (61), noted in a study of 2,205 patients, 655 of whom have osteoarthritis, that the WOMAC Index satisfied the requirements of Rasch item response theory across all disorders studied. Some items did not fit well, but it was not felt likely that removing them would be advantageous and might decrease usefulness in clinic and epidemiologic studies by restricting the range of the scale.

Predictive Value

Lingard et al (679), have reported that "low" pre-op WOMAC function scores are predictive of worse WOMAC function 12-months post total knee replacement ($p < 0.0001$). In a separate study by Ethgen et al (83) there was a 196% increase in rheumatologist visits in patients in the worst quartile for WOMAC function scores (cf best quartile) ($p < 0.05$).

SF-WOMAC

Several propositions for an SF-WOMAC have emerged(11,32,52,58). Four analyses have been based on all three WOMAC subscales and two have been based on only the function subscale. Sample sizes have varied from 224-1545 patients and studies have differed in environment (orthopaedic vs rheumatology), geography (European vs non-European), research design and analytic strategy. Based on a secondary analysis of the aforementioned studies, encompassing 4013 patients, we have noted differences between the different propositions. This disparity may be due to differences in clinical environment, geography, and methodology. While WOMAC questions #1-3, 5-7, 9, 10, 12, 13 and 18 have been identified in at least 50% of analyses, this exact configuration has not been part of any individual proposition. Given that all WOMAC items have been included in at least one proposition, the continued use of all 24 WOMAC questions is recommended. There are two potential negative consequences of short-forming: 1) Content validity may decline, and 2) Question deletion, based on data from one environment, may have a negative impact on index performance in other research and clinical environments. The WOMAC Index is brief and easily administered, and administration of the entire 24-item WOMAC questionnaire meets practicality requirements. These issues notwithstanding there may be situations where short forming is desirable and in those instances, the use of a short form, based on the aforementioned propositions, may be preferable (28).

Applications in non-OA disorders

Although not developed for non-OA applications, Hobby (1077) used the WOMAC Index in her MSc thesis involving 18 women with RA (functional class 2 and 3, mean age 59.2 years, mean disease duration 11.1 years, and which involved aerobic exercise and a test battery including a maximal treadmill stress test, the Functional Status Index, Arthritis Impact Measurement Scale and the WOMAC Index, and observed statistically significant improvements in WOMAC scores post- versus pre-training. No significant differences were detected by the FSI or AIMS indices. This work may suggest an application for the WOMAC Index in non-OA disorders of the lower extremity. Further research is recommended.

ADMINISTRATIVE ISSUES

Since its development, we have had an opportunity to use the WOMAC Index under a variety of conditions. A number of different versions of the Index have been created. **It is strongly recommended that users only employ original authorised versions of the WOMAC Index.**

We have provided support to a large number of registered users and addressed their questions regarding administration of the Index. The paragraphs that follow contain information which we have found useful in meeting the needs of WOMAC users.

Formatting the Index

Osteoarthritis is a very different condition from rheumatoid arthritis, and consideration should be given to the formatting of disease-specific questionnaires. While one may wish to assess all affected joints in OA, it is often more useful to base the evaluation of a new treatment on a single anatomic region (most often the hips and/ or knees). Even with hip and knee OA, one is frequently confronted with either a single symptomatic affected joint or with bilateral disease in which one joint is more symptomatic than the other. For this reason, we consider it important when using the WOMAC Index, to decide whether one wishes the patient to report on questions regarding: 1) hips and knees in general, or 2) hips or knees, or 3) a specific joint (e.g. left knee or study knee or more severely affected hip, etc.). We have had experience administering different questionnaires to separately assess the symptoms of OA in the signal and non-signal knees in the same individual. The use of the term "most severely affected joint" is problematic and we do not recommend its use, since the most symptomatic joint may vary throughout the time course of the study. In general, for interventional studies, we prefer to specify the joint of interest, i.e., the signal or study joint.

Another issue previously alluded to is time frame. In general, we have instructed patients to answer WOMAC questions considering their symptoms over the preceding 48 hours. For long term studies, however, where assessment intervals exceed one month, the time frame could be lengthened to the last 2 weeks.

Finally, one has to decide whether to make prior scores available. Our experience, based on two studies, is that in practice it makes no difference. If prior scores are not to be provided, then the patient will complete a blank questionnaire at each assessment point, and the Index can be produced in booklet form or as part of the case report form (CRFs). If patients are to be shown their prior scores, then ideally one would print CRFs

with a separate page for each WOMAC item and provide space for serial responses over time to be scored on the same page (i.e., several rows of response scales applicable to a single question on each page).

Once the aforementioned formatting issues have been considered, the WOMAC booklets or CRFs can be printed. The questionnaire has been electronically type-set, and we have generally directed users to our local printer, or suggested they make arrangements in their own geographic location for questionnaires to be produced. It is important to discuss with the publisher the fact that the VA scales must be exactly 10 cm in length. It is also wise to check, not only the galley proofs, but also the final shipment for accuracy in the text, as well as in the accuracy of reproduction of the 10 cm VA scales. On occasion, we have received a batch that has been distorted even though previous batches had been accurate. When printing from electronic files the length may also vary unless the page set up is selected correctly.

Finally, scannable and computer based forms of the WOMAC Index are in development. These will facilitate scoring and recording, since data can be quickly transcribed and transmitted.

Index Presentation

The WOMAC Index is self-administered and does not require the presence of an interviewer. We have occasionally mailed the questionnaire to patients who could not attend our clinic, and did not encounter any difficulty in their completing either the LK 3.0 or VA 3.0 versions at home. However, the availability of an assistant offers several advantages. In particular, the assistant can review the patient instructions which are provided at the beginning of the questionnaire. They can also check that the patient has completed all questions before they leave the clinic. Finally, we have prepared a short document which explains the meaning of questions in the WOMAC inventory in order to clarify any ambiguities which may arise from time to time (Appendix II). We, ourselves, have not found it necessary to refer to this document, but it was requested several years ago by one investigator conducting a study in a multicultural population. We recommend that the assistant, if present, avoid influencing the patient's scoring of the questionnaire. The WOMAC Index has been designed and developed to be self-administered, without reference to any third party. Under usual conditions, the Index can be completed in approximately five minutes. We have recent experience with telephone administration of the WOMAC 3.0 index. In a cross-over study comparing telephone vs office completion of the WOMAC LK3.0 (16), and have noted only small between-method variation in scores: pain = 0.9%, stiffness = 2.6%, physical function = 2.6%, total WOMAC LK3.0 = 2.3%. We conclude that the WOMAC LK3.0 remains valid when administered by telephone.

Score Calculation

The first step in calculation is to take the data off the raw questionnaire. For LK 3.1 numerical values are assigned to each of the five response categories (0=none, 1=mild, 2=moderate, 3=severe, 4=extreme). For each WOMAC dimension, a subscale score is calculated by simple summation of the assigned values scored on component items. Thus, the range of possible subscale scores for the three dimensions is as follows: pain=0-20, stiffness = 0-8, physical function=0-68. [For

convenience these scores can be normalised, and expressed on 0-10 scales (or 0-4 scales)].

For VA 3.1 a ruler is used to determine the distance in millimetres from the left end marker of each analogue to the point at which the patient's mark intersects the horizontal line on the analogue. For each WOMAC dimension, a subscale score is calculated by simple summation of the assigned values scored on component items. Thus, the range of possible subscale scores for the three dimensions is as follows: pain=0-500, stiffness=0-200, physical function=0-1700.

For NRS 3.1 numerical values have been pre-assigned to response categories between 0 (none) and 10 (extreme).

For each WOMAC dimension, a subscale score is calculated by simple summation of the assigned values scored on component items. Thus, the range of possible subscale scores for the three dimensions is as follows: pain = 0-50, stiffness = 0-20, physical function = 0-170

For convenience all aforementioned scores can be normalised and expressed on 0-10 or 0-100 scales. As noted previously, we recommend that, in general, the dimensions be kept separate and the analysis conducted on a subscale-by-subscale basis. If it is necessary to combine the three subscale scores, we recommend readers review the section on "Weighting and Aggregation".

Occasionally patients place their mark outside designated areas. Fortunately, this occurs relatively infrequently. We suggest the following scoring rules. If the patient places a mark between the response boxes in the LK 3.0 Index, we attribute the mark to the closest box. If it is placed exactly between two boxes, we attribute it, by convention, to the higher category. If the mark is placed to the left of the "none" box, it scores 0, and if it is to the right of the "extreme" box, it scores 4. We treat placements outside of the end markers of the VA scale in a similar fashion (i.e., outside the left end = 0, outside the right end = 100). We prefer patients to use an "X" or "slash" mark that intersects the analogue between the end markers. Sometimes, however, patients will use a "√ ". We read the "√" from the point of inflection. If the mark is placed above or below the analogue, such that it does not make a discreet intersect, we drop a perpendicular, using a set square, from the intersect of the "X" or the inflection part of the "√ " or the midpart of the "slash" mark to make a proper intersect. The score is determined by measurement from the left end marker to the point of intersection.

Occasionally patients fail to complete all questions. This should not occur if the investigator, or assistant, checks the CRFs for completeness prior to the patient leaving the clinic. If, however, there are missing values, we suggest the following. If ≥ two pain, both stiffness, or ≥four physical function items are omitted, the patient's response is regarded as invalid and the deficient subscale(s) should not be used in the analysis. Where one pain, one stiffness, or 1-3 physical function items are missing, we suggest substituting the average value for the subscale in lieu of the missing item value(s). This method is similar to that employed for some other indices.

Statistical Issues

Parameters used in sample size calculation are often difficult to find. Indeed, such parameters vary for different types of studies, e.g. the variance estimates may differ between populations and anatomical areas of interest. Likewise, the minimum clinically important difference may differ depending

on: the trial design (washout- retreatment versus continuous therapy, or placebo-controlled versus active-controlled trials), the intervention (surgical versus pharmacologic versus physiotherapy), the anatomic area of interest (e.g. hips versus knees), and the patient population (allcomers versus the elderly). It is recommended that WOMAC users review our publications (1-31,34,36, 37,39,40,47-56), and those of other investigators, who have published studies using the WOMAC Index (32,33,35,38,41-46,57-1041) in order to identify variance estimates in groups of patients having similar characteristics to those in any newly proposed study. Investigators requiring further information should contact our office at the Centre of National Research on Disability and Rehabilitation Medicine (see Foreword) for up-to-date information on parameters referable to defined study populations, and for advice regarding sample size calculation, inclusion/exclusion criteria and responder criteria.

With respect to data analysis, we recommend that the data be tested for normality and a decision made whether to use parametric versus non-parametric statistical methods. Several different approaches to analysis are possible with WOMAC data. Advice regarding data analysis can be provided, and is best considered a priori.

Responder Criteria

In addition to traditional subscale-by-subscale analysis or Total WOMAC score-based analysis of group data, consideration might be given to individualised analysis using responder criteria, either as change criteria or state attainment criteria. The OARSI and OMERACT-OARSI responder criteria, provide a new consensus-based approach to response category assignment (1056,1068). The criteria are based, in part, on WOMAC data, and are applicable to studies using outcomes, such as the WOMAC Index and patient global assessments (2,8). There appears to be convergence between response status assignment based on expert opinion, OARSI and OMERACT-OARSI criteria and patient perception (24).

We have considered the minimum perceptible clinical improvement (MPCI) for the WOMAC Index based on rofecoxib studies and have proposed the following values: WOMAC Pain = 9.7 nu, WOMAC Stiffness = 10.0 nu, WOMAC function = 9.3 nu (37). With MPCI criteria we were able detect clinically important and statistically significant differences in knee OA patients treated with Hylan G-F20 + Appropriate Care vs those treated with Appropriate Care alone (7).

An alternate approach has been taken by Tubach et al who have proposed definitions for Minimal Clinically Important Improvement (MCII), based on pain (visual analog scale) and function (WOMAC subscale), in knee and hip osteoarthritis patients (56). The proposed absolute (and relative) values for MCII for knee and hip osteoarthritis are as follows: 1) -19.9 mm (-40.8%) and -15.3 mm (-32.0%) for VA pain, 2) - 18.3 mm (-39.0%) and -15.2 mm (-32.6%) for patient's global assessment 3) -9.1 (-26.0%) and -7.9 (-21.1%) for WOMAC function subscale score (56).

In addition we have considered not only the minimum perceptible improvement but also proposed a hierarchy of thresholds based on WOMAC 20%, 50% and 70% improvement criteria. Parallel criteria consider percentage improvements in pain alone (WOMAC 20P, WOMAC 50P,

WOMAC 70P), or in pain and either stiffness or function (WOMAC 20PFS, WOMAC 50PFS, WOMAC 70PFS) (6,9,). With these criteria we have been able to detect clinically important and statistically significant differences in knee OA patients treated with Hylan G-F20 + Appropriate Care vs those treated with Appropriate Care alone (6,9). We have subsequently confirmed the success of the WOMAC 20 50 70 approach in comparing outcomes of hip and knee OA patients during treatment with rofecoxib and ibuprofen vs placebo (23).

In contrast to defining threshold values for minimum levels of improvement that are important to patients, a second paradigm considers the health state attained, rather than the absolute or relative change achieved. This follows the philosophy that "better is good, but good is best". Tubach et al (53,54) have proposed definitions of Minimum Clinically Acceptable State (MCAS) values for hip and knee OA as follows: Pain VAS 36 mm (hip), 33 mm (knee) and WOMAC function subscale 35 points (hip), 33 points (knee). These same investigators (55) have also proposed a definition for Patient Acceptable Symptom State (PASS) values as follows: Pain VAS 35 mm (hip), 32 mm (knee) and WOMAC function subscale 34 points (hip), 31 points (knee). These publications notwithstanding, exact definition of what constitutes acceptable symptom severity to patients with OA remains to be established, across a broad spectrum of individuals, disease severity, environments, cultures, languages, and instruments. The inter-subject and intra-subject variability also require further elaboration. Nevertheless, and in the absence of normative data for the general and OA populations, the aforementioned values represent initial efforts to understand and quantify the complexities of state attainment. We have had favourable experience recently in using state attainment criteria based on the WOMAC Index to differentiate between groups of patients. The Bellamy et al Low Intensity Symptom Severity (BLISS) Index categorises patients according to the velocity, magnitude and durability of the therapeutic response, based on time (a-time to first being in the state, b-ever being in the state, c-number of visits or percentage of time in the state, and d- in the state at study completion) and magnitude (≤25nu, ≤20nu, ≤15nu, ≤10nu, and ≤5nu) (10). These low intensity symptom severity states, reflect the extent to which patients achieve a good state rather than an improved state of health. Using the aforementioned approach we have been able to detect clinically important and statistically significant differences in knee OA patients treated with Hylan G-F20 + Appropriate Care vs those treated with Appropriate Care alone (10). We have subsequently confirmed the success of the BLISS Index approach in comparing outcomes of hip and knee OA patients during treatment with rofecoxib and ibuprofen vs placebo (22).

The extent to which WOMAC 20 50 70 responder criteria (6,9,23) and the BLISS Index (10,22) approach find applicability requires further evaluation. However, high level improvements (e.g. WOMAC 70%), and BLISS states of ≤5nu are potentially attainable in clinical practice, and in clinical trials may be statistically detectable in appropriately powered studies (6,9,10,22,23). Indeed the capacity to calculate the number needed to treat (NNT) from clinical trials data, may in the future permit, from a composite of NNT

values, the derivation of an Index of Therapeutic Benefit (ITB).

Outcome assessment in osteoarthritis clinical trials is dependent on the use of valid, reliable, and responsive measurement procedures (1042-1073). The WOMAC Index has demonstrated fulfilment of these criteria, and is increasingly finding application in diverse clinical and research environments in different countries.

REFERENCES

Studies of the Clinimetric Properties of the WOMAC Index.

1. *Barr S, Bellamy N, Buchanan WW, et al. A comparative study of signal versus aggregate methods of outcome measurement based on the WOMAC Osteoarthritis Index. The Journal of Rheumatology. 1994;21:2106-2112.*

2. *Bell MJ, Bellamy N, Goldsmith CH, Tugwell P, Torrance GW, Raynauld J-P, Polisson R, Pericak D and Walker V. Evaluation of the OARSI and OMERACT-OARSI set of responder criteria in patients treated with hylan G-F 20 for knee osteoarthritis. Arthritis and Rheumatism. 2003;48(9)(Suppl):S482.*

3. *Bellamy N and Buchanan WW. A preliminary evaluation of the dimensionality and clinical importance of pain and disability in osteoarthritis of the hip and knee. Clinical Rheumatology. 1986;5(2):231-241.*

4. *Bellamy N and Buchanan WW. Outcome measurement in osteoarthritis clinical trials: the case for standardisation. Clinical Rheumatology. 1984;3(3):293-303.*

5. *Bellamy N, Bell MJ, Goldsmith CH, Torrance GW, Raynauld J, Tugwell P, Polisson R, Pericak D and Walker V. Examining minimal perceptible clinical improvement in patients treated with hylan G-F 20 for knee osteoarthritis. Ann Rheum Dis 2004;63 (Suppl 1) 357.*

6. *Bellamy N, Bell MJ, Goldsmith CH, Torrance GW, Raynauld J, Tugwell P, Polisson R, Pericak D and Walker V. WOMAC 20, 50, 70 response levels in patients treated with hylan G-F 20 for knee osteoarthritis. EULAR 2004, June 9-12, Berlin.*

7. *Bellamy N, Bell MJ, Goldsmith CH, Torrance GW, Raynauld J, Tugwell P, Polisson R, Pericak D and Walker V. Examining minimal perceptible clinical improvement criteria in patients treated with hylan G-F 20 for knee osteoarthritis. EULAR 2004, June 9-12, Berlin.*

8. *Bellamy N, Bell MJ, Goldsmith CH, Tugwell P, Torrance GW, Raynauld J-P, Polisson R, Pericak D and Walker V. Evaluation of the OARSI and OMERACT-OARSI sets of responder criteria in patients treated with hylan G-F 20 for knee osteoarthritis. steoarthritis and Cartilage. 2003;11(Suppl A):S80.*

9. *Bellamy N, Bell MJ, Goldsmith CH, Tugwell P, Torrance GW, Raynauld JP, Polisson R, Pericak D and Walker V. WOMAC 20, 50, 70 response levels in patients treated with hylan G-F 20 for knee osteoarthritis. Osteoarthritis and Cartilage. 2003;11(Suppl A):S82.*

10. *Bellamy N, Bell MJ, Goldsmith CH, Tugwell P, Torrance GW, Raynauld JP, Polisson R, Pericak D and Walker V. Symptom free response using the WOMAC pain score in patients treated with hylan G-F 20 for knee osteoarthritis. Osteoarthritis and Cartilage. 2003;11(Suppl):S42-3.*

11. *Bellamy N, Bolognese J, Barlas S, Sullivan E, Watson D, Straus W. Evaluation of factors influencing the item content of the Short-Form WOMAC VA3.1 (SF-WOMAC VA3.1) Osteoarthritis Index. Arthritis and Rheumatism. 2002;46(9)(Suppl):S114-5.*

12. *Bellamy N, Buchanan WW, Chalmers A, et al. A multicenter study of tenoxicam and diclofenac in patients with osteoarthritis of the knee. The Journal of Rheumatology. 1993;20:999-1004.*

13. *Bellamy N, Buchanan WW, Goldsmith CH, Campbell J and Duku E. Signal measurement strategies: Are they feasible and do they offer any advantage in outcome measurement in osteoarthritis? Arthritis and Rheumatism. 1990;33(5):739-745.*

14. *Bellamy N, Buchanan WW, Goldsmith CH, Campbell J and Stitt L. Validation study of WOMAC: a health status instrument for measuring clinically-important patient-relevant outcomes following total hip or knee arthroplasty in osteoarthritis. Journal of Orthopaedic Rheumatology. 1988;1:95-108.*

15. *Bellamy N, Buchanan WW, Goldsmith CH, Campbell J and Stitt LW. Validation study of WOMAC: A health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. The Journal of Rheumatology. 1988;15:1833-1840.*

16. Bellamy N, Campbell J, Hill J and Band P. *A comparative study of telephone versus onsite completion of the WOMAC 3.0 Osteoarthritis Index. The Journal of Rheumatology. 2002;29(4):783-786.*

17. *Bellamy N, Campbell J, Hill J, Goldman A and Band P. Validation of telephone-administration of the WOMAC Osteoarthritis Index. Annals of the Rheumatic Diseases Abstracts 1999, XIV EULAR Congress, Glasgow, June 6-11, 1999, p.256.*

18. *Bellamy N, Campbell J, Stevens J, Pilch L, Stewart C and Mahmood Z. Validation study of a computerized version of the WOMAC VA3.0 Osteoarthritis Index. Australian Rheumatology Association 38th Annual Scientific Conference, May 22-24, 1994, Melbourne. Scientific Programme.*

19. *Bellamy N, Campbell J, Stevens J, Pilch L, Stewart C and Mahmood Z. Validation study of a computerized version of the Western Ontario and McMaster Universities VA 3.0 Osteoarthritis Index. The Journal of Rheumatology. 1997;24:2413-2415.*

20. *Bellamy N, Goldsmith CH, Buchanan WW, Campbell J and Duku E. Prior score availability: Observations using the WOMAC Osteoarthritis Index. British Journal of Rheumatology. 1991;30:150-151.*

21. *Bellamy N, Kean WF, Buchanan WW, Gerecz-Simon E and Campbell J. Double blind randomized controlled trial of sodium meclofenamate (Meclomen) and diclofenac sodium (Voltaren): Post validation reapplication of the WOMAC Osteoarthritis Index. The Journal of Rheumatology. 1992;19:153-159.*

22. *Bellamy N, Krupa D, Markind JE, Reicin AS and Malmstrom K. Prevalence of patients with low intensity pain symptom severity states during treatment with rofecoxib and ibuprofen for osteoarthritis. Osteoarthritis Cart 2004 (Abstract) (In Press).*

23. *Bellamy N, Krupa D, Markind JE, Reicin AS and Malmstrom K. WOMAC 20, 50, 70 response levels in patients treated with rofecoxib for osteoarthritis. Osteoarthritis Cart. 2004 (Abstract) (In Press).*

24. *Bellamy N, Lybrand SG and Gee T. An evaluation of the convergence between three different methods of response status assignment (RSA) based on the WOMAC Osteoarthritis Index. Arthritis and Rheumatism. 2002;46(9)(Suppl):S159.*

25. *Bellamy N, Wells G and Campbell J. Relationship between severity and clinical importance of symptoms in osteoarthritis. Clinical Rheumatology. 1991;10(2):138-143.*

26. *Bellamy N, Wells GA and Campbell J. PARIS Sectogram: a method for weighting and aggregating the WOMAC Osteoarthritis Index. Second International Congress of the Osteoarthritis Research Society, December 9-11, 1994, Orlando, Florida. Osteoarthritis and Cartilage. 1994;2(Suppl 1):37.*

27. *Bellamy N.* Pain assessment in osteoarthritis: experience with the WOMAC osteoarthritis index. Seminars in Arthritis and Rheumatism. 1989;18(4 Suppl 2):14-17.

28. *Bellamy N. Observations on alternate propositions for a short-form WOMAC Osteoarthritis Index (SF-WOMAC). EULAR, Lisbon, Portugal, June 18-21, 2003.*

29. *Bellamy N. Osteoarthritis - An evaluative index for clinical trials. MSc Thesis. McMaster University, Hamilton, Canada. 1982.*

30. *Bellamy N. WOMAC: A 20-year experiential review of a patient-centered self-reported health status questionnaire. The Journal of Rheumatology. 2002;29(12):2473-2476.*

31. *Bischoff HA, Vondechend M, Bellamy N and Theiler R. Validation of a computer touch screen version of the WOMAC 3.1 Likert scale osteoarthritis questionnaire. Osteoarthritis and Cartilage. 2003;11(Suppl A):S75.*

32. *Conaghan PG, Slade A, Emery P and Tennant A. The development of a short-form WOMAC instrument. Arthritis and Rheumatism. 2002;46(9)(Suppl):S156.*

33. *Conaghan PG, Slade An, Rocca R, Emery P and Tennant A. A rasch analysis of the total WOMAC Osteoarthritis Index score.* Arthritis and Rheumatism. 2001;44(9)(Suppl):S185.

34. *Davies G, Ehrich E, Watson D, Bolognese J, Seidenberg B and Bellamy N. Minimal perceptible clinical improvement (MPCI) in OA: WOMAC questionnaire and global assessments. Annals of the Rheumatic Diseases Abstracts 1999, XIV EULAR Congress, Glasgow, June 6-11, 1999, p.255-256.*

35. *Davis AM, Badley EM, Beaton DE, Kopec J, Wright JG, Young NL and Williams JI. Rash analysis of the Western Ontario McMaster (WOMAC) Osteoarthritis Index: results from community and arthroplasty samples. Journal of Clinical Epidemiology. 2003;56:1076-1083.*

36. *Ehrich E, Davies G, Watson D, Bolognese J, Seidenberg B and Bellamy N. Minimal perceptible clinical improvement (MPCI) in osteoarthritis: WOMAC and global assessments, influence of age, gender and geographic region. Arthritis and Rheumatism. 1999;42(9)(Suppl):S143.*

37. *Ehrich EW, Davies GM, Watson DJ, Bolognese JA, Seidenberg BC and Bellamy N. Minimal perceptible clinical improvement with the Western Ontario and McMaster Universities Osteoarthritis Index questionnaire and global assessments in patients with osteoarthritis. The Journal of Rheumatology. 2000;27(11):2635-2641.*

38. *Falissard B, Ravaud P, Tubach F, Logeart I, Baron G and Dougados M. Determination of a minimal clinically meaningful difference in the WOMAC physical function score: application of a new concept, the MDP75. EULAR, Lisbon, Portugal, June 18-21, 2003.*

39. *Griffiths G, Bellamy N, Bailey WH, Bailey SI, McLaren AC and Campbell J. A comparative study of the relative efficiency of the WOMAC, AIMS and HAQ instruments in evaluating the outcome of total knee arthroplasty. Inflammopharmacology. 1995;3:1-6.*

40. *Griffiths G, Bellamy N, Kean WF, Campbell J and Gerecz-Simon E. A study of the time frame dependency of responses to the WOMAC Osteoarthritis Index. Inflammopharmacology. 1993;2:85-87.*

41. *Kennedy D, Stratford PW, Pagura SMC, Wessel J, Gollish JD and Woodhouse LJ. Exploring the factorial validity and clinical interpretability of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC). Physiotherapy Canada. 2003;55:160-168.*

42. *Lustig SP, Shaw AR, Cappelleri JC and Vazquez-Abad M-D. The WOMAC discriminates between patients with mild and severe osteoarthritis. Arthritis and Rheumatism. 2000;43(9)(Suppl):S110.*

43. *McConnell S, Kolopack P and Davis AM. The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC): a review of its utility and measurement properties. Arthritis Care and Research. 2001;45(5):453-461.*

44. *Pagura SMC, Stratford PW, Kennedy DM, Gollish JD and Woodhouse IJ. Improving the clinical utility of the Western Ontario McMaster Osteoarthritis Index: new findings. Physiotherapy Canada. 2002;54, no.1 (Supplement):S14.*

45. Ryser L, Wright BD, Aeschlimann A, Marlacher-Gehler S and Stucki G. A new look at the Western Ontario and McMaster *Universities Osteoarthritis Index using Rasch analysis. Arthritis Care and Research. 1999;12(5):331-335.*

46. *Sataffi F, Carotti M, Stancati A and Grassi W. Quality of life in patients with osteoarthritis of the hip or the knee: comparison of generic and disease-specific instruments. EULAR 2004, June 9-12, Berlin, Germany.*

47. *Theiler R, Bellamy N, Kroesen S and Uebelhart D. Multisite validation study of the WOMAC 3.1 LK OA index using a computerized touchscreen version in a phase IV rofecoxib trial. Osteoarthritis and Cartilage. 2001;9(Suppl B):S46.*

48. *Theiler R, Kroesen S, Spielberger J and Bellamy N. Validation of the WOMAC 3.0 TS, an audiovisual computerized version of the WOMAC 3.0. Osteoarthritis and Cartilage. 2000;8(2):152-153.*

49. *Theiler R, Spielberger J, Bischoff HA, Bellamy N, Huber J, Kroesen S. Clinical evaluation of the WOMAC 3.0 OA Index in numeric rating scale format using a computerized touch screen version. Osteoarthritis and Cartilage. 2002;10:479-481.*

50. *Tubach F, Baron G, Falissard B, Logeart I, Dougados M, Bellamy N and Ravaud P. Development of a French short form of the WOMAC function scale based on the patients' and experts' perspective. Osteoarthritis and Cartilage. 2003;11(Suppl A):S37-8.*

51. *Tubach F, Baron G, Falissard B, Logeart I, Dougados M, Bellamy N and Ravaud P. Development of a French short form of the WOMAC function scale based on the patients' and experts' perspective. EULAR, Lisbon, Portugal, June 18-21, 2003.*

52. *Tubach F, Baron G, Falissard B, Logeart I, Dougados M, Bellamy N and Ravaud P. Development of a French short form of the WOMAC function scale based on the patients' and experts' perspective. Arthritis and Rheumatism. 2003;48(9)(Suppl):S485.*

53. *Tubach F, Baron G, Ravaud P, Falissard B, Logeart I, Bombardier C, Felson D, Van der Heijde D, Hochberg M, Bellamy N and Dougados M. Evaluation of clinically relevant changes and states in symptomatic outcome variables in osteoarthritis. Arthritis and Rheumatism. 2003;48(9)(Suppl):S292.*

54. *Tubach F, Baron G, Ravaud P, Falissard B, Logeart I, Bombardier C, Felson D, Van der Heijde D, Hochberg M, Bellamy N and Dougados M. Evaluation of clinically relevant changes and states in symptomatic outcome variables in osteoarthritis. EULAR, Lisbon, Portugal, June 18-21, 2003.*

55. *Tubach F, Ravaud P, Baron G, Falissard B, Logeart I, Bellamy N, Bombardier C, Felson DT, Hochberg MC, van der Heijde D and Dougados M. Evaluation of clinically relevant states in patient-reported outcomes in knee and hip osteoarthritis: the Patient Acceptable Symptom State. Ann Rheum Dis. Online publication 18 June 2004. doi:10.1136/ard.2004.023028.*

56. *Tubach F, Ravaud P, Baron G, Falissard B, Logeart I, Bellamy N, Bombardier C, Felson DT, Hochberg MC, van der Heijde D and Dougados M. Evaluation of clinically relevant changes in patient-reported outcomes in knee and hip osteoarthritis: the Minimal Clinically Important Improvement. Ann Rheum Dis. Online publication 18 June, 2004. doi:10.1136/ard.2004.022905.*

57. *Villanueva I, Toyos FJ, Ariza-Ariza R and Navarro F. Relative efficiency and validity properties of a visual analogue (VA) versus a Likert (LK) scaled version of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC). Arthritis and Rheumatism. 1999;42(9)(Suppl):S147.*

58. *Whitehouse SL, Learmonth ID, Lingard EA & K.P.O.S. Group [Bristol]. A reduced WOMAC function scale – derivation and validity. The Journal of Bone and Joint Surgery [Br]. 2003;85-B(Suppl I):17.*

59. *Whitehouse SL, Lingard EA, Katz JN and Learmonth ID. Development and testing of a reduced WOMAC function scale. The Journal of Bone and Joint Surgery. 2003;85-B(5):706-711.*

60. *Whitehouse SL, Lingard EA, Learmonth ID and Kinemax Outcomes Group. A reduced WOMAC function scale - derivation and validation. Arthritis and Rheumatism. 2000;43(9)(Suppl.):S393.*

61. *Wolfe F and Kong SX. Rasch analysis of the Western Ontario McMaster Questionnaire (WOMAC) in 2205 patients with osteoarthritis, rheumatoid arthritis, and fibromyalgia. Annals of the Rheumatic Diseases. 1999;58:563-568.*

62. *Wolfe F and Kong SX. Rasch item response theory applied to the WOMAC in osteoarthritis. Rheumatology in Europe. 1998;26(Suppl 2):121.*

63. *Wolfe F. Back pain rather than radiographic abnormality explains WOMAC pain and function scores in patients with osteoarthritis and other rheumatic disorders. Arthritis and Rheumatism.1998;41(9)(Suppl):S145.*

64. *Wolfe F. Determinants of WOMAC function, pain and stiffness scores: evidence for the role of low back pain, symptom counts, fatigue and depression in osteoarthritis, rheumatoid arthritis and fibromyalgia. Rheumatology. 1999;38:355-361.*

Comparative Studies of WOMAC versus Health Related Quality of Life (HRQOL) Measures.

65. *Ang DC, Ibrahim SA, Burant CJ and Kwoh CK. Is there a difference in the perception of symptoms between African Americans and whites with osteoarthritis? The Journal of Rheumatology. 2003;30(6):1305-1310.*

66. *Angst F, Aeschlimann A and Stucki G. Smallest detectable and minimal clinically important differences of rehabilitation intervention with their implications for required sample sizes using WOMAC and SF-36 quality of life measurement instruments in patients with osteoarthritis of the lower extremities. Arthritis Care and Research. 2001;45:384-391.*

67. *Angst F, Aeschlimann A, Michel BA and Stucki G. Minimal clinically important rehabilitation effects in patients with osteoarthritis of the lower extremities. The Journal of Rheumatology. 2002;29:131-138.*

68. *Angst F, Aeschlimann A, Steiner W and Stucki G. Responsiveness of the WOMAC osteoarthritis index as compared with the SF-36 in patients with osteoarthritis of the legs undergoing a comprehensive rehabilitation intervention. Annals of the Rheumatic Diseases. 2001;60:834-840.*

69. *Bachmeier CJM, March LM, Cross M, et al. A comparison of WOMAC and MOS SF-36 in OA patients undergoing joint replacement. Australian Rheumatology Association. 40th Annual Scientific Conference, Brisbane 25-28, May 1997. Scientific Programme. p.87.*

70. *Bálint G, Lovas K, Szende A, Hodinka L, Héjj G and Szebenyi B. Quality of life and burden of disease in osteoarthritis. A survey of 150 patients in Hungary. Annals of the Rheumatic Diseases. 2000;59(Suppl 1).*

71. *Beaton D, Erdeljan S, Hogg-Johnson S and Bombardier C. Comparison of performance of generic SF-36 and 3 disease specific measures of health status in injured workers. Arthritis and Rheumatism. 1995;38(9)(Suppl):S175.*

72. *Bolognese J, Ehrich E, Schnitzer T and Zeng Q. Precision of composite measures of osteoarthritis efficacy compared with that of individual endpoints. Arthritis and Rheumatism. 1998;41(9)(Suppl):S361.*

73. *Bolognese JA, Schnitzer TJ and Ehrich EW. Response relationship of VAS and Likert scales in osteoarthritis efficacy measurement. Osteoarthritis and Cartilage. 2003;11:499-507.*

74. *Bombardier C, Melfi CA, Paul J, et al. Comparison of a generic and a disease-specific measure of pain and physical function after knee replacement surgery. Medical Care. 1995;33(4)(Suppl):AS131-144.*

75. *Bombardier C, Melfi CA, Paul J, Green R and Hawker G. Comparison of a generic and a disease-specific measure of pain and physical function after knee replacement surgery. Arthritis and Rheumatism. 1994;37(9)(Suppl):S225.*

76. *Brazier JE, Harper R, Munro J, Walters SJ and Snaith ML. Generic and condition-specific outcome measures for people with osteoarthritis of the knee. Rheumatology. 1999;38:870-877.*

77. *Brooks P, Oh E-S, Cross M, et al. A comparison of WOMAC and MOS SF-36 in OA patients undergoing joint replacement. Arthritis and Rheumatism. 1996;39(9)(Suppl):S259.*

78. *Brooks PM, Bachmeier CJM, March LM, et al. A comparison of WOMAC and MOS SF-36 in OA patients undergoing joint replacement. Arthritis and Rheumatism. 1997;40(9)(Suppl):S110.*

79. *Creamer P, Lethbridge-Cejku M and Hochberg MC. A model of the health effects of osteoarthritis (OA) of the knee. Arthritis and Rheumatism. 1998;41(9)(Suppl):S229.*

80. *Davies GM, Watson DJ and Bellamy N. Comparison of the responsiveness and relative effect size of the Western Ontario and McMaster Universities Osteoarthritis Index and the Short-Form Medical Outcomes Study Survey in a randomized, clinical trial of osteoarthritis patients. Arthritis Care and Research. 1999;12(3):172-179.*

81. *Dedhiya SD, Zhao SZ, Bocanegra TS, et al. Effect on health related quality of life among patients with osteoarthritis of the knee treated with oxaprozin potassium. The Journal of Rheumatology. 1998;25(Suppl 52):5.*

82. *Diaz-Borjon E, Marcial-Barba LD, Cardiel MH, Hernandez-Cruz BE, Ruiz-Mercado R and Alcocer-Varela J. Changes in quality of life in rheumatic patients with joint replacement. Comparison of generic versus specific instruments. Arthritis and Rheumatism. 1995;38(9)(Suppl):S165.*

83. *Ethgen O, Kahler KH, Kong SX, Reginster J-Y and Wolfe F. Are health-related quality of life scores useful in predicting the use of health care services? Exploration in patients with arthritis. Arthritis and Rheumatism. 2000;43(9)(Suppl):S163.*

84. *Ethgen O, Kahler KH, Kong SX, Reginster J-Y and Wolfe F. Health-related quality of life scores and future use of health care resources in patients with rheumatoid arthritis and osteoarthritis: Implication for health services research in arthritis. Annals of the Rheumatic Diseases. 2000;59(Suppl 1).*

85. *Ethgen O, Kahler KH, Kong SX, Reginster J-Y and Wolfe F. The effect of health related quality of life on reported use of health care resources in patients with osteoarthritis and rheumatoid arthritis: a longitudinal analysis. The Journal of Rheumatology. 2002;29:1147-1155.*

86. *Ethgen O, Van Parijs P, Delhalle S, Rosant S, Bruyere O and Reginster JY. Importance of social companionship in the determination of health-related quality of life in hip and knee osteoarthritis. Osteoarthritis and Cartilage. 2001;9(Suppl B):S19-20.*

87. Fransen M and Edmonds J. Validity of EUROQOL in patients with osteoarthritis of the knee. Australian Rheumatology Association. 41st Annual Scientific Conference, Sydney 24-27, May 1998. Scientific Programme. p.157.

88. Fransen M and Edmonds J. Reliability and validity of the EuroQol in patients with osteoarthritis of the knee. Rheumatology. 1999;38:807-813.

89. Fransen M, Salkeld G and Edmonds J. Validity of the EuroQol in patients with osteoarthritis of the knee. Arthritis and Rheumatism. 1998;41(9)(Suppl):S224.

90. Geusens P, Manicourt D, Masure J and Diepenbeek LK. Quality of life in patients with osteoarthritis treated with non-steroidal anti-inflammatory drugs - a case-controlled survey. Arthritis and Rheumatism. 1999;42(9)(Suppl):S294.

91. Green Gwinn SM, Starz TW, Rogers J, Chandler L and Vogt M. Disease specific or general health status measures in knee osteoarthritis: which characterizes function best? Arthritis and Rheumatism. 2001;44(9)(Suppl):S165.

92. Hawker G, Melfi C, Paul J, Green R and Bombardier C. Comparison of a generic (SF-36) and a disease specific (WOMAC) (Western Ontario and McMaster Universities Osteoarthritis Index) instrument in the measurement of outcomes after knee replacement surgery. The Journal of Rheumatology. 1995;22:1193-1196.

93. Ibrahim SA, Burant CJ, Siminoff LA and Kwoh K. Perceptions of quality of life in patients with chronic knee or hip pain: differences between African-American and white veterans. Arthritis and Rheumatism. 1999;42(9)(Suppl):S72.

94. Ibrahim SA, Siminoff LA, Burant CJ and Kwoh CK. Variation in perceptions of treatment and self-care practices in elderly with osteoarthritis: A comparison between African American and white patients. Arthritis Care and Research. 2001;45:340-345.

95. Jogi P. Comparison of the WOMAC and the LEFS questionnaires in patients awaiting and having undergone total knee arthroplasty. Master of Science Thesis, School of Physical Therapy, The University of Western Ontario, London, Canada, April 2003, 66 pages.

96. Jones JG and Leighton F. Comparison of WOMAC with SF-36 and OA of the knee or hip. Annals of the Rheumatic Diseases. 2002;61:182-183.

97. Kirschner S, Walther M, Böhm D, Matzer M, Heesen T, Faller H and König A. German short musculoskeletal function assessment questionnaire (SMFA-D): comparison with the SF-36 and WOMAC in a prospective evaluation in patients with primary osteoarthritis undergoing total knee arthroplasty. Rheumatology International. 2003;23(1):15-20.

98. König A, Kirschner S, Walther M, Böhm D, Matzer M, Heesen T and Faller H. Cultural adaptation and evaluation of the short Musculoskeletal Function Assessment Questionnaire in patients with osteoarthritis of the knee. American Academy of Orthopaedic Surgeons 68th Annual Meeting, Feb. 28-Mar. 4, 2001, San Francisco, U.S.A. (Paper 065).

99. Lapsley HM, March LM, Tribe KL, Cross MJ and Brooks PM. Living with osteoarthritis: patient expenditures, health status, and social impact. Arthritis Care and Research. 2001;45:301-306.

100. Lothbridge-Cejku M, Creamer P and Hochberg M. Quality of life in knee osteoarthritis (OA). Arthritis and Rheumatism. 1997;40(9)(Suppl):S174.

101. Lethbridge-Cejku M, Creamer P, Hochberg MC, Zeng Q, Bolognese J and Ehrich E. The health effects of hip osteoarthritis. Arthritis and Rheumatism. 2000;43(9)(Suppl):S340.

102. March LM, Lapsley H, Cross M, Byrne E, Tribe K and Brooks P. Long term out of pocket expenditure and health outcomes for subjects living with osteoarthritis (OA) and rheumatoid arthritis (RA): a 9-year cohort study. EULAR 2004, June 9-12, Berlin, Germany.

103. March LM, Oh E-S, Cross M, et al. A comparison of WOMAC and MOS SF-36 in OA patients undergoing joint replacement. 8th APLAR Congress of Rheumatology. April 21-26, 1996, Melbourne, Australia. Programme and Abstracts. Abstract #439 (Posters P119).

104. Martin DP, Engelberg R, Agel J and Swiontkowski MF. Comparison of the Musculoskeletal Function Assessment questionnaire with the Short Form-36, the Western Ontario and McMaster Universities Osteoarthritis Index, and the Sickness Impact Profile Health-Status measures. The Journal of Bone and Joint Surgery [Am]. 1997;79-A(9):1323-1335.

105. McCarthy CJ and Oldham JA. The reliability, validity and responsiveness of an aggregated locomotor function (ALF) score in patients with osteoarthritis of the knee. Rheumatology. 2004;43(4):514-517.

106. Müller ME. Creating benchmarks of the Standard Evaluation Questionnaire (SEQ) for disorders of the locomotor system. A nationwide project to improve quality assessment in health care. Center for Continuing Education and Documentation in Orthopaedic Surgery [MEM-CED]. University of Berne, Switzerland. 8 pages.

107. Pouchot J, Fautrel B, Maigret P and Meric G. Satisfaction of osteoarthritis (OA) patients with their treatments: results of a national survey among hip or knee OA patients cared by general practitioners (GPS). EULAR 2004, June 9-12, Berlin, Germany.

108. Sangha O, Thoilor R, Schären S and Stucki G. Comparative responsiveness of health status instruments and clinical variables in osteoarthritis (OA). Arthritis and Rheumatism. 1997;40(9)(Suppl):S235.

109. Segal L, Osborne RH, Day S and Chapman A. The priority setting in osteoarthritis project: evidence-based health sector wide modeling of health outcomes, economics, and quality of life. Osteoarthritis and Cartilage. 2002;10(Suppl A):S72.

110. Theiler R, Schären St, Sangha O, Dick W, Tyndall A and Stucki G. Comparative responsiveness of general and condition-specific health status measures in osteoarthritis (OA). Rheumatology in Europe. 1997;26(Suppl 2):81.

111. Tubach F, Ravaud P, Baron G, Falissard B, Logeart I, Bellamy N, Bombardier C, Felson D, Hochberg M, van der Heijde D and Dougados M. Evaluation of clinically relevant changes and states in symptomatic outcome variables in osteoarthritis of the lower limbs. Osteoarthritis and Cartilage. 2003;11(Suppl A): S35.

112. Van der Waal JM, Terwee CB, Van der Windt DAWM, Bouter LM and Dekker J. The impact of chronic hip and knee complaints in general practice on health-related and overall quality of life. EULAR 2004, June 9-12, Berlin, Germany.

113. Villanueva I, del Mar Guzman M, Javier Toyos F, Ariza R and Navarro F. Defining clinically important differences in knee osteoarthritis from the patient's perspective. Arthritis and Rheumatism. 2003;48(9)(Suppl):S294.

114. Weigl M, Cieza A, Harder M, Geyh S, Amann E, Kostanjsek N and Stucki G. Linking osteoarthritis-specific health-status measures to the International Classification of Functioning, Disability, and Health (ICF). Osteoarthritis and Cartilage. 2003;11(7):519-523.

115. Wolfe F, Kong SX and Watson DJ. Effect of disease duration on the quality of life of patients with arthritis. The Journal of Rheumatology. 1998;25(Suppl 52):65.

116. Wolfe F, Kong SX and Watson DJ. Quality of life and its non-arthritis predictors in rheumatoid arthritis, osteoarthritis and fibromyalgia: A comparative study of 2,463 patients. The Journal of Rheumatology. 1998;25(Suppl 52):65.

117. Wright JG and Young NL. A comparison of different indices of responsiveness. Journal of Clinical Epidemiology. 1997;50(3):239-246.

Application of WOMAC in Studies of Pharmacologic Agents, Viscosupplementation and Physical Forms of Therapy.

118. Acevedo E, Romanowicz A, Van der Heijde D, Bolognese J, Seldenberg B, Ehrich E, Mukhopadhyay S, Daniels B for the Rofecoxib Multinational Protocol 034 Study Group. Rofecoxib, a COX-2 specific inhibitor (C-2,SI), had clinical efficacy comparable to diclofenac in the treatment of knee and hip osteoarthritis (OA) in a one-year controlled clinical trial. Annals of the Rheumatic Diseases Abstracts 1999. XIV EULAR Congress, Glasgow, June 6-11, 1999, p.206.

119. Allen E, Yates A, Walczyk J, Buttles S, Berkowitz J and Krohn K. Valgus bracing for isolated medial compartment osteoarthritis of the knee. Arthritis and Rheumatism. 2000;43(12):2861.

120. Altman R, Moskowitz R and the Hyalgan Study Group. Intraarticular sodium hyaluronate (Hyalgan) in the treatment of patients with osteoarthritis of the knee: A randomized clinical trial. The Journal of Rheumatology. 1998;25:2203-2212.

121. Altman RD and Marcussen KC. Effects of a ginger extract on knee pain in patients with osteoarthritis. Arthritis and Rheumatism. 2001;44(11):2531-2538.

122. Altman RD and the IAP Study Group. Ibuprofen, acetaminophen and placebo in osteoarthritis of the knee: a six-day double-blind study. Arthritis and Rheumatism. 1999;42(9)(Suppl):S403.

123. Altman RD, Marcusson KC and Christensen FN. Study of a ginger/galanga extract in osteoarthritis (OA) of the knee. Arthritis and Rheumatism. 2000;43(9)(Suppl):S341.

124. Altman RD, Marcussen KC, Sigwart K and Christensen FN. Extract of zingiber officinale (ginger) and alpinia galanga in osteoarthritis. Annals of the Rheumatic Diseases. 2000;59(Suppl 1).

125. Arboleya LR, de la Figuera E, Soledad Garcia M, Aragón B and the VICOXX Study Group. Experience of rofecoxib in patients with osteoarthritis previously treated with traditional non-steroidal anti-inflammatory drugs in Spain: results of phase 2 of the VIOXX study. Current Medical Research and Opinion. 2003;19(4):288-297.

126. Ardic F, Bolulu D, Topuz O and Cubukcu S. Efficacy of intra-articular hyaluronic acid injections in knee osteoarthritis. Annals of the Rheumatic Diseases. 2001;60(Suppl I):232.

127. Arokoski MH, Arokoski JPA, Haara M, Vesterinen M, Niemitukia L and Helminen HJ. Muscle function and hip muscle cross-sectional area in men with and without hip osteoarthritis. Osteoarthritis and Cartilage. 2002;10(Suppl A):S64.

128. Atamaz F, Kirazli Y and Akkoc Y. A comparative study between intra-articular hylan G-F 20 and Na-hyaluronate and physical therapy in the management of knee osteoarthritis. EULAR 2004, June 9-12, Berlin.

129. Ates A, Kinikli G, Turgay M and Duman M. Viscosupplementation with sodium hyaluronate for the treatment of osteoarthritis. The Journal of Rheumatology. 2001;28(Suppl 63):8 (ILAR, Edmonton, 2001).

130. Auerbach B and Melzer C. Die Behandlung der Gonarthrose mit hochvernetzter Hyaluronsäure – Ergebnisse einer prospektiven randomisierten Studie (Cross-linked hyaluronic acid in the treatment of osteoarthritis of the knee – results of a prospective randomized trial). Zentralblatt für Chirurgie. 2002;127:895-899.

131. Auerbach B. Ergebnisse einer prospektiven randomisierten Studie zur Wirksamkeit hochvernetzter Hyaluronsäure bei der Behandlung der Gonarthrose. Inaugural-Dissertation. Giessen 2001.

132. *Ayral X, Macklilop N, Genant HK, Kirkpatrick J, Beaulieu A, Pipplngskiöld P, Will RK, Alava S and Dougados M. Arthroscopic evaluation of potential structure-modifying drug in osteoarthritis of the knee. A multicenter, randomized, double-blind comparison of tenidap sodium vs piroxicam. Osteoarthrits and Cartilage. 2003;11(3):198-207.*

133. *Babul N, Noveck R, Chipman HN, Roth SH, Gana T, Albert K and the Biovail Tramadol ER Osteoarthritis Group. A double blind randomized 12 week placebo controlled trial of tramadol ER in osteoarthritis of the knee. Arthritis and Rheumatism. 2002;46(9)(Suppl):S157.*

134. *Bae S-C, Kim T-H, Jun J-B, Jung S-S, Yoo D-H and Kim S-Y. Non-cognitive decline in the elderly osteoarthritis patients with COX-2 specific inhibitor: a pilot study. Arthritis and Rheumatism. 2001;44(9)(Suppl):S71.*

135. *Bagga H, Burkhardt D, Morris D, Ghosh P and March L. Effect of intra-articular hylan GF-20 (Synvisc) on synovial fluid and clinical parameters. Osteoarthritis and Cartilage. 2002;10(Suppl A):S69-S70.*

136. *Bagga H, Burkhardt D, Morris D, Ghosh P, Sambrook P and March L. Long term effects of intra-articular hylan GF-20 (Synvisc) on synovial fluid hyaluronan and sulfated glycosaminoglycans levels. EULAR, Lisbon, Portugal, June 18-21, 2003.*

137. *Baker K, Nelson M, Layne J, Felson D and Roubenoff R. The efficacy of home-based progressive strength training for knee osteoarthritis (KOA). Arthritis and Rheumatism. 1999;42(9)(Suppl):S292.*

138. *Baker K, Niu J, Goggins J, Clancy M and Felson D. The effects of vitamin C intake on pain in knee osteoarthritis (OA). Arthritis and Rheumatism. 2003;48(9)(Suppl):S422.*

139. *Baker KR, Nelson ME, Felson DT, Layne JE, Sarno R and Roubenoff R. The efficacy of home based progressive strength training in older adults with knee osteoarthritis: a randomized controlled trial. The Journal of Rheumatology. 2001;28:1655-1665.*

140. *Bell M, Fallaha M, Lenczner E, Ranger P and Welsh P. Viscosupplementation with Hylan G-F 20 in total knee replacement candidates: an effective pain management therapy that may delay surgery. Osteoarthritis and Cartilage. 1999;7(Supplement A):p.S30.*

141. *Bellamy N, Torrance G, Raynauld JP, Goldsmith C, Tugwell P, Schultz M, Bourne RB, Marshall KW, Band P and Walker V. A randomized controlled trial evaluating effectiveness and cost-effectiveness of Hylan G-F 20 in patients with knee osteoarthritis. American Academy of Orthpaedic Surgeons 68th Annual Meeting, Feb. 28-Mar. 4, 2001, San Francisco, U.S.A. (Paper 072).*

142. *Bellamy N, Torrance G, Raynauld JP, Goldsmith C, Tugwell P, Schultz M, Walker V, Bourne RB, Marshall KW and Band P. A randomized controlled trial evaluating effectiveness and cost-effectiveness of Hylan G-F 20 in patients with knee osteoarthritis. American Academy of Orthopaedic Surgeons 68th Annual Meeting, Feb. 28-Mar. 4, 2001, San Francisco, U.S.A. (Scientific Exhibitor SE04).*

143. *Bellamy N, Yocum D, Roszko P and Hall D. Pharmacodynamic dissection of dose-response to meloxicam in osteoarthritis using the WOMAC index. The Journal of Rheumatology. 2001;28(Suppl 63):6 (ILAR, Edmonton, 2001).*

144. *Benevolenskaya L, Tüzün S, Hagin E, Moore A and Gimona A. Lumiracoxib is effective in relieving symptoms of knee or hip osteoarthritis after 4 weeks of treatment: results from a randomized, placebo-controlled trial. EULAR, Lisbon, Portugal, June 18-23, 2003.*

145. *Bensen WG, Fiechtner JJ, McMillen JI, et al. Treatment of osteoarthritis with celecoxib, a cyclooxygenase-2 inhibitor: a randomized controlled trial. Mayo Clin Proc. 1999; 74:1095-1105.*

146. *Berenbaum F, Zacher J, Fleischmann R, Moore A and Gimona A. Efficacy of lumiracoxib in osteoarthritis demonstrated using the newly developed OMERACT-OARSI criteria. EULAR 2004, June 9-12, Berlin, Germany.*

147. *Berg P and Olsson U. Intra-articular injection of non-animal stabilised hyaluronic acid (NASHA) for osteoarthritis of the hip. Osteoarthritis and Cartilage. 2003;11(Suppl A):S45.*

148. *Berman BM, Lao L, Greene M, et al. Efficacy of traditional Chinese acupuncture in the treatment of symptomatic knee osteoarthritis: a pilot study. Osteoarthritis and Cartilage. 1995;3:139-142.*

149. *Berman BM, Singh BB, Lao L, Langenberg P, Li H, Hadhazy V, Bareta J and Hochberg M. A randomized trial of acupuncture as an adjunctive therapy in osteoarthritis of the knee. Rheumatology. 1999;38:346-354.*

150. *Bias P and Buchner A. Licofelone – a 5-LOX, COX-1and COX-2 inhibitor – is as effective as naproxen and has improved GI tolerability in patients with osteoarthritis. Arthritis and Rheumatism. 2003;48(9)(Suppl):S72.*

151. *Bilgici A, Akdeniz O, Kuru O and Ulusoy H. The effect of a home-based exercise therapy versus an aerobic exercise programme on pain and functional disability in patients with knee osteoarthritis. EULAR 2004, June 9-12, Berlin, Germany.*

152. *Binette M, Yates AJ Jr, Krohn K, Walczyk J and Berkowitz J. Valgus bracing for isolated medial compartment osteoarthritis of the knee. American Academy of Orthopaedic Surgeons 68th Annual Meeting, Feb. 28-Mar. 4, 2001, San Francisco, U.S.A. (Poster PE143).*

153. *Birbara C, Sheldon E, Farrell J, Rodgers A, Petruschke R, Robertson D and Chang DJ. Comparable efficacy of rofecoxib 12.5 mg and celecoxib 200 mg in the treatment of patients with osteoarthritis: a randomized, controlled trial. EULAR 2004, June 9-12, Berlin, Germany.*

154. Blanco FJ, Bias P, Buchner A, Lammerich A and Schulz U. Over 52 weeks, licofelone – a 5-lipoxygenase, cyclooxgenase-1 and -2 inhibitor – is as effective as naproxen and shows improved tolerability in osteoarthritis. Osteoarthritis and Cartilage. 2003;11(Suppl A):S40.

155. Blanco FJ, Buchner A, Bias P, Lammerich A and Schulz U. Licofolone, an inhibitor of COX-1, COX-2 and 5-LOX, is as effective as naproxen and shows improved safety during 12 months of treatment in patients with osteoarthritis of the knee. EULAR, Lisbon, Portugal, June 18-21, 2003.

156. Blanco-Garcia FJ, Fernandez-Sueiro JL, Pinto-Tasende JA, Fernandez-Lopez JC, Ramos M, Gimeno M, Coronel P and Galdo F. Intraarticular hyaluronic acid (Adant) in the treatment of patients with knee osteoarthritis waiting by replacement surgery. EULAR 2004, June 9-12, Berlin.

157. Bocanegra TS, Kivitz A, Hurley SM, et al. Treatment of osteoarthritis (OA) with diclofenac/misoprostol is associated with a lower incidence of gastroduodenal ulcers (GDU) compared to nabumetone and naproxen. Arthritis and Rheumatism. 1998; 41(9)(Suppl):S197.

158. Bolgen Cimen O, Gulaldi NCM, Ozisik S, Bagis S, Kanik A and Erdogan C. Clinical and scintigraphic evaluation in intra-articular hyaluronate therapy. EULAR, Lisbon, Portugal, June 18-21, 2003.

159. Bookman A, Shainhouse JZ and Williams K. A multicenter study of diclofenac topical lotion (Pennsaid) in patients with osteoarthritis of the knee. Osteoarthritis and Cartilage. 1999;7(Supplement A): p.S31.

160. Bookman AAM, Shainhouse JZ and Williams K. A multicenter study of diclofenac topical lotion (Pennsaid) in patients with osteoarthritis of the knee. Arthritis and Rheumatism. 1999;42(9)(Suppl):S142.

161. Brand C, Snaddon J, Bailey M and Cicuttini F. Vitamin E is ineffective for symptomatic relief of knee osteoarthritis: a six month double-blind, randomised, placebo controlled study. Annals of the Rheumatic Diseases. 2001;60:946-949.

162. Brandt KD, Block JA, Michalski JP, Moreland LW, Caldwell JR, Lavin PT and the ORTHOVISC Study Group. Efficacy and safety of intraarticular sodium hyaluronate in knee osteoarthritis. Clinical Orthopaedics and Related Research. 2001;No.385:130-143.

163. Brandt KD, Heilman DK, Slemenda C, Katz BP, Mazzura S, Braunstein EM and Byrd D. A comparison of lower extremity muscle strength, obesity, and depression scores in elderly subjects with knee pain with and without radiographic evidence of knee osteoarthritis. The Journal of Rheumatology. 2000;27(8):1937-1946.

164. Brandt KD, Mazzuca SA, Katz BP, Lane KA and the OA-Doxycycline Study Group. The disease-modifying effect of doxycycline (Doxy) includes symptomatic benefit for patients with knee osteoarthritis. Arthritis and Rheumatism. 2003;48(12):3653.

165. Brown DJ and Beinat L. Safety and efficacy of an hyaluronan of 500-730 KDa and hylan G-F 20 in clinical practice. Osteoarthritis and Cartilage. 2003;11(Suppl A):S118.

166. Bruce B, Singh G, Sato R and Waddell DD. Hylan G-F 20 significantly improves pain, function and patient global after 6 months in knee OA. American Academy of Orthopaedic Surgeons 71st Annual Meeting, March 10-14, 2004, San Francisco, Paper Presentation No. 159.

167. Bruyere O, Compere S, Rovati LC, Giacovelli G, Deroisy R and Reginster JY. Five-year follow-up of patients from a previous 3-year randomised, controlled trial of glucosamine sulfate in knee osteoarthritis. Osteoarthritis and Cartilage. 2003;11(Suppl A):S10-11.

168. Bruyere O, Compere S, Rovati LC, Giacovelli G, Deroisy R and Reginster J-Y. Five-year follow-up of patients from a previous 3-year randomised, controlled trial of glucosamine sulfate in knee osteoarthritis. Arthritis and Rheumatism. 2003;48(9)(Suppl):S80.

169. Bruyere O, Honore A, Ethgen O, Rovati LC, Giacovelli G, Henrotin YE, Seidel L and Reginster JYL. Correlation between radiographic severity of knee osteoarthritis and future disease progression. Results from a 3-year prospective, placebo-controlled study evaluating the effect of glucosamine sulfate. Osteoarthritis and Cartilage. 2003;11(1):1-5.

170. Buchner A, Bias P and Schulz U. Licofelone, a 5-lox, cox-1 and cox-2 inhibitor, is as effective as celecoxib over 12 weeks of treatment: OARSI/OMERACT-OARSI response criteria results. EULAR 2004, June 9-12, Berlin, Germany.

171. Burch F, Codding C, Patel N and Sheldon E. Lidocaine patch 5% improves pain, stiffness, and physical function in osteoarthritis pain patients. A prospective, multicenter, open-label effectiveness trial. Osteoarthritis and Cartilage. 2004;12:253-255.

172. Caborn D, Rush J, Lanzer W, Parenti D, Murray C on behalf of the Synvisc 901 Study Group. A randomized, single-blind comparison of the efficacy and tolerability of hylan G-F 20 and triamcinolone hexacetonide in patients with osteoarthritis of the knee. The Journal of Rheumatology. 2004;31(2):333-343.

173. Caborn DN. Efficacy and tolerability of hylan G-F 20 compared to intra-articular triamcinolone hexacetonide in patients with osteoarthritis of the knee in a randomized, evaluator-blinded study. EULAR, Lisbon, Portugal, June 18-21, 2003.

174. Caldwell JR, Rapoport RJ, Davis JC, Offenberg HL, Marker HW, Roth SH, Yuan W, Eliot L, Babul N, Mikkelsen Lynch P for the Avinza TRG004-04 Study Group. Efficacy and safety of a once-daily morphine formulation in chronic, moderate-to-severe osteoarthritis pain: results from a randomized, placebo-controlled, double-blind trial and an open-label extension trial. Journal of Pain and Symptom Management. 2002;23(4):278-291.

175. Campbell R, Quilty B and Dieppe P. Discrepancies between patients' assessments of outcome: qualitative study nested within a randomised controlled trial. BMJ. 2003;326:252-253.

176. Campbell R, Quilty B and Dieppe P. What constitutes change in response to interventions for osteoarthritis? Osteoarthritis and Cartilage. 2001;9(Suppl B):S26-27.

177. Cannon G, Caldwell J, Holt P, et al. MK-0966, a specific COX-2 inhibitor, has clinical efficacy comparable to diclofenac in the treatment of knee and hip osteoarthritis (OA) in a 26-week controlled clinical trial. Arthritis and Rheumatism. 1998;41(9)(Suppl):S196.

178. Cannon GW and Breedveld FC. Efficacy of cyclooxygenase-2-specific inhibitors. The American Journal of Medicine. 2001; 110(3A):6S-12S.

179. Cannon GW, Caldwell JR, Holt P, et al. Rofecoxib, a specific inhibitor of cyclooxygenase 2, with clinical efficacy comparable with that of diclofenac sodium. Arthritis and Rheumatism. 2000;43(5):978-987.

180. Case JP and Block JA. Pain relief in osteoarthritis (OA) of the knee. Osteoarthritis and Cartilage. 2001;9(Suppl B):S23-24.

181. Case JP, Baliunas AJ and Block JA. Lack of efficacy of acetaminophen in treating symptomatic knee osteoarthritis. A randomized, double-blind, placebo-controlled comparison trial with diclofenac sodium. Archives of Internal Medicine. 2003;163:169-178.

182. Cheras PA, Nielsen GVL, Blackwell EA, Outerbridge HK and Fine S. Role of vascular and haematological abnormalities in osteoarthritis (OA) - rationale for use of pentosan polysulphate and glucosamine/chrondroitin sulphate. Osteoarthritis and Cartilage. 2001;9(Suppl B):S44.

183. Chevalier X, Conrozier T, Bertin P, Bailleul F and Vignon E. Treatment of patients with symptomatic hip OA - a pilot study with intra-articular hylan G-F20 (Synvisc). Arthritis and Rheumatism. 2000;43(9)(Suppl):S384.

184. Chhabra SR, Newlove R and Montgomery S. Intraarticular hylan injections in the treatment of osteoarthritis of the knee: a prospective study of 120 patients. Journal of Bone and Joint Surgery [Br]. 2000;82-B(Suppl. 1):p.68.

185. Childs JD, Sparto PJ, Fitzgerald GK and Bizzini M. The relationship between alterations in lower extremity movement and muscle activation patterns in individuals with knee osteoarthritis and self-report and performance-based measures of function and disability. Journal of Orthopaedic & Sports Physical Therapy. 2003;33(2):A-11.

186. Chivukula L, Sulochana S and Hameed H. Rofecoxib relieves symptoms and controls progression of OA knee. The Journal of Rheumatology. 2001;28(Suppl 63):8 (ILAR Edmonton 2001).

187. Chopra A, Lavin P, Chitre D, Patwardhan B and Polisson R. A clinical study of an Ayurvedic (Asian Indian) medicine in OA knees. Arthritis and Rheumatism. 1998;41(9)(Suppl):S198.

188. Cibere J, Esdaile JM, Thorne A, Singer J, Kopec JA, Canvin J, Robinson DB, Pope J, Hong P, Grant E, Ionescu M and Poole AR. Multicenter randomized double-blind placebo-controlled glucosamine discontinuation trial in osteoarthritis. Arthritis and Rheumatism. 2002;46(9)(Supplement):S576-577.

189. Clark DI, Downing N, Mitchell J, Coulson L, Syzpryt EP and Doherty M. Physiotherapy for anterior knee pain: a randomised controlled trial. Annals of the Rheumatic Diseases. 2000;59:700-704.

190. Clarke S, Lock V, Duddy J, Newman J and Kirwan JR. Intra-articular hylan GF20 (Synvisc) in the management of patellofemoral osteoarthritis of the knee (POAK). Annals of the Rheumatic Diseases. 2001;60(Suppl I):230-231.

191. Clemett D and Goa KL. Celecoxib. A review of its use in osteoarthritis, rheumatoid arthritis and acute pain. Drugs. 2000;59(4):957-980.

192. Cohen M, Wolfe R and Mai T. A randomised, double-blind, placebo-controlled trial of topical glucosamine and chondroitin sulfate for osteoarthritis of the knee. Internal Medicine Journal. 2003;33(Suppl):A17.

193. Cohen M, Wolfe R, Mai T and Lewis D. A randomized, double blind, placebo controlled trial of a topical cream containing glucosamine sulfate, chondroitin sulfate, and camphor for osteoarthritis of the knee. The Journal of Rheumatology. 2003;30:523-528. [Correction: J Rheumatol 2003;30(11):2512]

194. Cohen M, Wolfe R, Mai T and Lewis D. Topical glucosamine and chondroitin sulfate for osteoarthritis of the knee: a randomized controlled trial. Osteoarthritis and Cartilage. 2002;10(Suppl A):S4.

195. Cohen MA, Shiroky JB, Ballachey M-L, Neville C and Esdaile JM. Double-blind randomized trial of intra-articular (I/A) hyaluronate in the treatment of osteoarthritis of the knee. Arthritis and Rheumatism. 1994;37(6)(Suppl):R31.

196. Cohen SB, Fleischmann RM, Schnitzer TH and Vitow C. Double blind, controlled dose ranging study of tepoxalin in knee osteoarthritis. Arthritis and Rheumatism. 1994;37(9)(Suppl):S241.

197. Colker CM, Swain M, Lynch L and Gingerich DA. Effects of a milk-based bioactive micronutrient beverage on pain symptoms and activity of adults with osteoarthritis: a double-blind, placebo-controlled clinical evaluation. Nutrition. 2002;18(5):388-392.

198. Conaghan PG, Day RO and Brooks P for the Australian Multicentre Trial Group. Beneficial effects of intraarticular hyaluronan (ARTZ) in osteoarthritis of the knee - results of a multicentre study. Australian Rheumatology Association. 40th Annual Scientific Conference, Brisbane 25-28, May 1997. Scientific Programme. p.89.

199. Conaghan PG, Keenan A-M, Duffin PA, Hampshire MJ, Rankine JJ and Robinson P. The efficacy of intra-articular corticosteroids for osteoarthritis of the hip. Arthritis and Rheumatism. 2003;48(9)(Suppl):S76.

200. Conrozier T, Bertin P, Bailleul F and Chevalier X. Efficacy and safety of Hylan GF-20 (Synvisc) in patients with hip osteoarthritis: preliminary results of a multicentre pilot study. Annals of the Rheumatic Diseases. 2000;59(Suppl 1).

201. Conrozier T, Bertin P, Bailleul F, Mathieu P, Trèves R, Vignon E and Chevalier X. Intra-articular injection of hylan GF-20 (Synvisc) in patients with hip osteoarthritis: preliminary results of efficacy and safety in a multicentre pilot study. Osteoarthritis and Cartilage. 2000;8(Suppl B):S66-67.

202. Conrozier T, Bertin P, Mathieu P, Charlot J, Bailleul F, Treves R, Vignon E and Chevalier X. Intra-articular injections of hylan G-F 20 in patients with symptomatic hip osteoarthritis: An open-label, multicentre, pilot study. Clinical and Experimental Rheumatology. 2003;21:605-610.

203. Creamer P, Singh BB, Hochberg MC and Berman BM. Are psychosocial factors related to response to acupuncture among patients with knee osteoarthritis? Alternative Therapies in Health and Medicine. 1999;5(4):72-76.

204. Curtis SP, De Tora LM, Leung AT, Sanchez MM, Connors L and Ng J. Etoricoxib showed rapid onset of efficacy and sustained duration of therapy over the 24-hour dosing interval in patients with osteoarthritis (OA) of the knee or hip. Annals of the Rheumatic Diseases. 2002; 61 (Supplement). EULAR, June 12-15, Stockholm.

205. Curtis SP, Fisher C, Kafka S, Bochow B, Ko A, Compton A, Pramanik B and Gertz BJ. Treatment with etoricoxib (MK-0663), a COX-2 selective inhibitor, resulted in clinical improvement in knee osteoarthritis (OA) over 52 weeks. Annals of the Rheumatic Diseases. 2001;60(Suppl I):229-230.

206. Curtis SP, Krupa DA, Ko AT, Bockow B, Fisher C, Compton A, Mehta A, Geissler L, De Tora L and Reicin A. Etoricoxib improves pain, stiffness and physical function in OA patients. EULAR, Lisbon, Portugal, June 18-21, 2003.

207. Curtis SP, Leung A, Watson DJ, Ramey, Kong S and Harper S. Improved quality-of-life (QoL) with etoricoxib compared with placebo in patients with osteoarthritis. Osteoarthritis and Cartilage. 2001;9(Suppl B):S24.

208. Daniels B, Krupa D, Ehrich E, Bolognese J and Seidenberg B. Clinical response of OA patients who use acetaminophen when randomized to rofecoxib or ibuprofen. Arthritis and Rheumatism. 1999;42(9)(Suppl):S143.

209. Das A Jr and Hammad TA. Efficacy of a combination of FCHG49 glucosamine hydrochloride, TRH122 low molecular weight sodium chondroitin sulfate and manganese ascorbate in the management of knee osteoarthritis. Osteoarthritis and Cartilage. 2000;8(5):343-350.

210. Das SK, Mishra K, Ramakrishnan S, Srivastava R, Agarwa GG, Singh R and Sircar AR. A randomized controlled trial to evaluate the slow-acting symptom modifying effects of a regimen containing colchicine in a subset of patients with osteoarthritis of the knee. Osteoarthritis and Cartilage. 2002;10(4):247-252.

211. Das SK, Ramakrishnan S, Mishra K, Srivastava R, Agarwal GG, Singh R and Sircar AR. A randomized controlled trial to evaluate the slow-acting symptom-modifying effects of colchicine in osteoarthritis of the knee: a preliminary report. Arthritis Care & Research. 2002;47(3):280-284. DOI 10.1002/art.10455.

212. Day R, Brooks P, Conaghan PG, Petersen M for the Multicenter Trial Group. A double blind, randomized, multicenter, parallel group study of the effectiveness and tolerance of intraarticular hyaluronan in osteoarthritis of the knee. The Journal of Rheumatology. 2004;31(4):775-782.

213. Day R, Luza A, Casteneda O, Strusberg A, Bolognese J, Seidenberg B, Krupa D, Morrison B and Ehrich E. Rofecoxib, a cox-2 specific inhibitor (C-2 SI), had clinical efficacy comparable to ibuprofen in the treatment of knee and hip osteoarthritis (OA) in a 6-week controlled clinical trial. Annals of the Rheumatic Diseases Abstracts 1999. XIV EULAR Congress, Glasgow, June 6-11, 1999, p.206.

214. Day R, Morrison B, Luza A, Castaneda O, Strusberg A, Nahir M, Helgetveit KB, Kress B, Daniels B, Bolognese J, Krupa D, Seidenberg B, Ehrich E and the Rofecoxib/Ibuprofen comparator study group. A randomized trial of the efficacy and tolerability of the COX-2 inhibitor rofecoxib vs ibuprofen in patients with osteoarthritis. Archives Internal Medicine. 2000;160:1781-1787.

215. De Tora LM, Krupa D, Bolognese J, Sperling RS and Ehrich EW. Rofecoxib shows consistent efficacy in osteoarthritis clinical trials, regardless of specific patient demographic and disease factors. The Journal of Rheumatology. 2001;28(11):2494-2503.

216. De Tora LM, Krupa D, Bolognese JA, Sperling RS and Ehrich EW. Rofecoxib shows consistent efficacy in OA clinical trials, regardless of specific patient demographic and disease factors. Annals of the Rheumatic Diseases. 2001;60(Suppl I):231.

217. Deyle GD, Allison SC, Matekel RL, Ryder MG, Henderson NE, Hutton JP, Stang JM, Gohdes DD and Garber MB. Physical therapy treatment effectiveness for osteoarthritis of the knee: a comparison of supervised clinical exercise and manual therapy procoduros versus a home exercise program. Journal of Orthopaedic & Sports Physical Therapy. 2001;31(1):A3.

218. Deyle GD, Henderson NE, Matekel RL, Ryder MG, Garber MB and Allison SC. Effectiveness of manual physical therapy and exercise in osteoarthritis of the knee. A randomized, controlled trial. Annals of Internal Medicine. 2000;132(3):173-181.

219. Dickson DJ, Hosie G, English JR and the Primary Care Rheumatology Society OA Knee Study Group. A double-blind, placebo-controlled comparison of hylan G-F 20 against diclofenac in knee osteoarthritis. Journal of Drug Assessment. 2001;4:179-190 [First published as Journal of Clinical Research 2001;4:41-52.]

220. *Dickson J and Gillian Hosie on behalf of the Primary Care Rheumatology Society OA Knee Study Group. Double-blind, double-control comparison of viscosupplementation with hylan GF 20 (Synvisc) against diclofenac and control in knee osteoarthritis. Arthritis and Rheumatism. 1998;41(9)(Suppl):S197.*

221. *Dickson J, Hosie G on behalf of the Primary Care Rheumatology Society OA Knee Study Group. Double-blind, double-control comparison of viscosupplementation with Synvisc against diclofenac and control in knee osteoarthritis. British Journal of Rheumatology. 1998;37(Suppl 1):155.*

222. *Dominick KL, Bosworth HB, Hsieh JB and Moser BK. Perceptions of analgesic and anti-inflammatory medication helpfulness among patients with osteoarthritis. Arthritis and Rheumatism. 2003;48(9)(Suppl):S423.*

223. *Ehrich E, Schnitzer T, Kivitz A, et al. MK-966, a highly selective COX-2 inhibitor, was effective in the treatment of osteoarthritis (OA) of the knee and hip in a 6-week placebo controlled study. Arthritis and Rheumatism. 1997;40(9)(Suppl):S85.*

224. *Ehrich E, Schnitzer T, Weaver A, et al. Treatment with MK-966 (VIOXX), a specific COX-2 inhibitor, resulted in clinical improvement in osteoarthritis (OA) of the knee and hip, that was sustained over six months. Rheumatology in Europe. 1998;27(Suppl 2):p.119.*

225. *Ehrich E, Schnitzer TJ, McIlwain H, Levy R, Wolfe F, Weisman M, Zeng Q, Morrison B, Bolognese J, Seidenberg B, Gertz BJ for the Rofecoxib Osteoarthritis Pilot Study Group. Effect of specific COX-2 inhibition in osteoarthritis of the knee: a 6 week double blind, placebo controlled pilot study of rofecoxib. The Journal of Rheumatology. 1999;26:2438-2447.*

226. *Eskiyurt N. The first experience of the efficacy and safety of coxibs in Turkish nation. Annals of the Rheumatic Diseases. 2001;60(Suppl I):325.*

227. *Espinoza L, Cohen S, Moskowitz R, Bahrt K. Celecoxib in Hispanic-Americans: efficacy, safety and tolerability compared to Naproxen or placebo in the treatment of osteoarthritis of the knee. EULAR, Lisbon, Portugal, June 18-21, 2003.*

228. *Espinozoa L, Cohen S, Moskowitz R and Bahrt K. Celecoxib in Hispanic-Americans: efficacy, safety, and tolerability compared to naproxen or placebo in the treatment of osteoarthritis of the knee. EULAR, Lisbon, Portugal, June 18-21, 2003.*

229. *Evcik D and Sonel B. Effectiveness of a home-based exercise therapy and walking programe on osteoarthritis of the knee. Rheumatology International. 2002;22(3):103-106.*

230. *Evcik D and Sonel B. Effectiveness of home based exercise therapy and walking program on osteoarthritis of the knee. Annals of the Rheumatic Diseases. 2002; 61 (Supplement). EULAR, June 12-15, Stockholm.*

231. *Evcik D, Kuru I and Maralcan G. The effects of exercise therapy in patients with patello-femoral pain. EULAR 2004, June 9-12, Berlin, Germany.*

232. *Eyigor S, Hepguler S and Capaci K. A comparison of muscle training methods in patients with knee osteoarthritis. Clinial Rheumatology. 2004;23:109-115.*

233. *Eyigor S, Hepguler S and Capaci K. Isokinetic and progressive resistive exercises in the treatment of knee osteoarthritis. Arthritis and Rheumatism. 2001;44(9)(Suppl):S354.*

234. *Fiechtner JJ, Sikes D and Recker D. A double-blind, placebo-controlled dose ranging study to evaluate the efficacy of valdecoxib, a novel COX-2 specific inhibitor, in treating the signs and symptoms of osteoarthritis of the knee. Annals of the Rheumatic Diseases. 2001;60(Suppl I):56.*

235. *Fisher C, Bockow B, Curtis SP, Miller E, Olalcye J, Compton A, McCormick CL and Reicin AS. Results of a 190-week randomized study demonstrated similar efficacy of etoricoxib compared with diclofenac in patients with osteoarthritis (OA). EULAR, Lisbon, Portugal, June 18-21, 2003.*

236. *Fisher CA, Curtis SP, Resnick H, Ripley P, Leung AT, Ko AT, Bergman GA, De Tora LM and Gertz BJ. Treatment with etoricoxib, a COX-2 selective inhibitor, resulted in clinical improvement in knee and hip osteoarthritis (OA) over 52 weeks. Arthritis and Rheumatism. 2001;44(9)(Suppl):S135.*

237. *Fitzgerald GK, Childs JD, Ridge TM and Irrgang JJ. Agility and perturbation training for a physically active individual with knee osteoarthritis. Physical Therapy. 2002;82(4):372-382.*

238. *Fleischmann R, Sheldon E, Maldonado Cocco J, Yu S, Dutia D and Usiskin K. A prospective randomized 13-week study evaluating the efficacy of lumiracoxib in patients with osteoarthritis of the knee. EULAR, Lisbon, Portugal, June 18-23, 2003.*

239. *Fleischmann RM, Caldwell JR, Roth SH, Tesser JRP, Olson W and Kamin M. Tramadol for the treatment of joint pain associated with osteoarthritis: a randomized, double-blind, placebo-controlled trial. Current Therapeutic Research - Clinical and Experimental. 2001;62:113-128.*

240. *Fleischmann RM, Kamin M and Olson WH. Safety and efficacy of tramadol for the signs and symptoms of osteoarthritis. Arthritis and Rheumatism. 1999;42(9)(Suppl):S144.*

241. *Foley A, Halbert J, Hewitt T and Crotty M. Does hydrotherapy improve strength and physical function in patients with osteoarthritis – a randomised controlled trial comparing a gym based and a hydrotherapy based strengthening programme. Annals of the Rheumatic Diseases. 2003;62:1162-1167.*

242. *Forero JH, Alfonso R, Arabelaez F, Hunsche E and Kong SX. Evaluation of rofecoxib in patients with osteoarthritis under real-life conditions – the VALOR study in Colombia. EULAR, Lisbon, Portugal, June 18-21, 2003.*

243. *Forero JH, Alfonso R, Arabelaez F, Hunsche E and Kong SX. Real-life evaluation of rofecoxib in patients with osteoarthritis in Colombia – the VALOR study. Osteoarthritis and Cartilage. 2003;11(Suppl A):S37.*

244. *Fransen M, Crosbie J and Edmonds J. Muscle force in osteoarthritis of the knee. Osteoarthritis and Cartilage. 2002;10(Suppl A):S10.*

245. *Fransen M, Crosbie J and Edmonds J. Physical therapy is effective for patients with osteoarthritis of the knee: a randomized controlled clinical trial. The Journal of Rheumatology. 2001;28:156-164.*

246. *Fransen M, Crosbie J and Edmunds J. Predictors of responsiveness to physiotherapy treatment for patients with osteoarthritis of the knee. Osteoarthritis and Cartilage. 2000;8(Suppl B):S71.*

247. *Fransen M, Crosbie J and Edmunds J. Randomized controlled clinical trial: effectiveness of physiotherapy for patients with osteoarthritis of the knee. Osteoarthritis and Cartilage. 2000;8(Suppl B):S13.*

248. *Frizziero L, Graninger W, Langevitz P, Kortesuo H, Moser U, Rodger IW, Vandormael K and Ozturk ZE. Etoricoxib is effective in osteoarthritis patients with different levels of baseline severity of the disease. EULAR 2004, June 9-12, Berlin, Germany.*

249. *Gagnier P and Hwang E. Meloxicam versus usual care NSAIDS for the treatment of osteoarthritis in a usual care setting. The results of the IMPROVE trial. Annals of the Rheumatic Diseases. 2002; 61 (Supplement). EULAR, June 12-15, Stockholm.*

250. *Galer BS, Gammaitoni AR, Burch F, Codding C, Patel N et al. Impact of treatment with a 5% lidocaine patch on pain, stiffness, and physical function in patients with osteoarthritis. Archives of Physical Medicine and Rehabilitation. 2003;84(9):E22.*

251. *Geba GP, Polis AB, Dixon ME, Dobbins TW, Rush JE and Weir MR. Comparative blood pressure effects of rofecoxib, celecoxib, and placebo in patients with osteoarthritis (OA): a randomized controlled trial. Annals of the Rheumatic Diseases. 2001;60(Suppl I):238.*

252. *Geba GP, Schnitzer TJ, Polis AB, Petruschke RA and Weaver AL. Patient global assessment and WOMAC response among osteoarthritis patients treated with rofecoxib, celecoxib, or acetaminophen in the VACT-2 trial. EULAR, Lisbon, Portugal, June 18-21, 2003.*

253. *Geba GP, Weaver AL, Polis AB, Dixon ME, Schnitzer TJ for the VACT Group. Efficacy of rofecoxib, celecoxib, and acetaminophen in osteoarthritis of the knee. A randomized trial. JAMA. 2002;287(1):64-71.*

254. *Geba GP, Weaver AL, Polis AB, Petruschke RA and Schnitzer TJ. Analysis of the onset of efficacy with rofecoxib, celecoxib, and acetaminophen in osteoarthritis: the VACT-2 trial. EULAR, Lisbon, Portugal, June 18-21, 2003.*

255. *Geba GP, Weaver AL, Polis AB, Petruschke RA and Schnitzer TJ. The response of osteoarthritis patients to therapy with rofecoxib, celecoxib, or acetaminophen: a pooled analysis of the VACT trials. EULAR, Lisbon, Portugal, June 18-21, 2003.*

256. *Geba GP, Weaver AL, Schnitzer TJ, Fleischmann RM, Polis AB, Matzura-Wolfe DM, Dixon ME and Gormley GJ. A comparison of rofecoxib to celecoxib and acetaminophen in the treatment of osteoarthritis. Arthritis and Rheumatism. 2000;43(9)(Suppl):S384.*

257. *Geba GP, Weaver AL, Schnitzer TJ, Polis AB, Dixon ME, Kivitz AJ and Gormley GJ. A clinical trial comparing rofecoxib to celecoxib and acetaminophen in the treatment of osteoarthritis (OA): early efficacy results. Annals of the Rheumatic Diseases. 2000;59(Suppl 1).*

258. *Geis GS, Hubbard RC, Yu S and Zhao W. Efficacy and tolerability of celecoxib: a comparison of once-daily vs twice-daily dosing. Arthritis and Rheumatism. 1999; 42(9)(Suppl):S144.*

259. *Gerard D, Brocq O, Roux CH, Albert C and Euller-Ziegler L. Intra-articular injection of hyaluronic acid (Durolane) for hip osteoarthritis: An open study. EULAR, June 9-12, Berlin.*

260. *Gibofsky A, Fort JG, Mangal B, Wolfe F, Simon LS, Moskowitz RW, Zlotnick S and Lei H. Correlation between patient preference assessment and traditional measures of efficacy: results from 2 crossover, randomized, controlled trials comparing placebo, acetaminophen, and celecoxib (the PACES Trials) in patients with osteoarthritis. Arthritis and Rheumatism. 2003;48(9)(Suppl):S293.*

261. *Gibofsky A, Williams GW, McKenna F and Fort JG. Comparing the efficacy of cyclooxygenase 2-specific inhibitors in treating osteoarthritis. Appropriate trial design considerations and results of a randomized, placebo-controlled trial. Arthritis and Rheumatism. 2003;48(11):3102-3111.*

262. *Gibofsky A, Wolfe F, Simon L, Moskowitz R, Fort J, Zlotnick S, Mangal B and Lei H. Correlation between patient preference assessment and traditional measures of efficacy: results from a double-blind, cross-over trial comparing placebo, acetaminophen (paracetamol), and celecoxib effectiveness study (PACES-1) in patients with osteoarthritis. EULAR, Lisbon, Portugal, June 18-21, 2003.*

263. *Goldman RJ, Reinbold KA, Iglarsh ZA, Oatis C, Newstadter L, Zonay L and Schumacher HR. Osteo-arthritic knee isometric exerciser with feedback. Arthritis and Rheumatism. 1999;42(9)(Suppl):S287.*

264. Goldsmith CH, Raynauld JP, Torrance GW, Band PA, Bellamy N, Tugwell P, Walker V and Schultz M. A randomized controlled trial evaluating effectiveness and cost-effectiveness of hylan G-F 20 in patients with knee OA. The Journal of Rheumatology. 2001;28(Suppl 63):9 (ILAR, Edmonton, 2001).

265. Gottesdiener K, Schnitzer T, Fisher C, Bockow B, Markenson J, Ko A, De Tora L, Curtis S, Geissler L and Gertz BJ for the Protocol 007 Study Group. Results of a randomized, dose-ranging trial of etoricoxib in patients with osteoarthritis. Rheumatology. 2002;41:1052-1061.

266. Gottesdiener K, Schnitzer T, Fisher C, Kafka S, Bockow B, Dalgin P, Markenson J, Shahane A, Ng J, Gertz B for the MK-663 Osteoarthritis Phase II Study Group. MK-663, a specific cox-2 inhibitor for treatment of osteoarthritis (OA) of the knee. Arthritis and Rheumatism. 1999;42(9)(Suppl):S144.

267. Gottsdiener K, Schnitzer T, Fisher C, Kafka S, Bockow B, Dalgin P, Markenson J, Ko A and Gertz B. Treatment with MK-663, a specific COX-2 inhibitor, resulted in clinical improvement in osteoarthritis (OA) of the knee that was sustained over three months. Annals of the Rheumatic Diseases. 2000;59(Suppl 1).

268. Goupille P, Giraudeau B, Conrozier T, Marliere J, Kiefer P and Chevalier X. Safety and efficacy of intra-articular injection of IL-1ra (IL-1 receptor antagonist) in patients with painful osteoarthritis of the knee: a multicenter, double blind study. Arthritis and Rheumatism. 2003;48(9)(Suppl):S696-S697.

269. Grace D, Rogers J, Skeith K and Anderson K. Topical diclofenac versus placebo: a double blind, randomized clinical trial in patients with osteoarthritis of the knee. The Journal of Rheumatology. 1999;26:2659-2663.

270. Graff JS, Remmers A, Nyberg J, Martin SA, Wright T and Agin M. Assessment of physical functioning in osteoarthritis: a comparison of two assessments in knee or hip osteoarthritis patients. EULAR, Lisbon, Portugal, June 18-21, 2003.

271. Groppa L, Golubciuc S, Vremis L, Dutca L and Sincari L. The efficacy of combined condroprotective treatment in the osteoarthritis of the knee. EULAR 2004, June 9-12, Berlin, Germany.

272. Guerrero R, Herrero-Beaumont G, Acebes C, Sánchez-Pernaute O, Llorens MA, Blázquez AB, Palacios I, Egido J and Vivanco F. Cartilage turnover and symptomatic relief observed in patients with knee OA after intra-articular administration of a hyaluronate polymer (Adant). Osteoarthritis and Cartilage. 1999;7(Supplement A):p.S32.

273. Guerrero R, Sanchez-Pernaute O, Acebes C, Llorens MA, Mas S, Palacios I, Egido J and Herrero-Beaumont G. Modification of synovial cartilage markers after repetitive intraarticular administration of hyaluronate (Adant) in knee OA patients. Arthritis and Rheumatism. 1999;42(9)(Suppl):S253.

274. Guler M, Kuran B, Parlar D, Guler M, Saglam H, Yapici S, Guzeloglu S, Ozgul Messe F and Boneval F. Clinical trial of intra-articular injection of hyaluronic acid in patients with osteoarthritis of the knee. Poster presented at the X National Rheumatology Congress, October 29 - November 3, 1996, Pamukkale-Denizli, Turkey.

275. Gur A, Kylync S, Sarac AJ, Cevik R and Nas K. Comparison of effects of meloxicam and physiotherapy in the management of serious knee osteoarthritis. Annals of the Rheumatic Diseases. 2002; 61 (Supplement). EULAR, June 12-15, Stockholm.

276. Gur A, Kylync S, Sarac AJ, Nas K and Cevik R. Effects of intra-articular hyaluronic acid and corticosteroid in the patients with serious knee osteoarthritis. Annals of the Rheumatic Diseases. 2002; 61 (Supplement). EULAR, June 12-15, Stockholm.

277. Gur A, Sarac AJ, Kylync S, Cevik R and Nas K. Effects of intra-articular betamethasone and meloxicam in the patients with serious knee osteoarthritis. Annals of the Rheumatic Diseases. 2002; 61 (Supplement). EULAR, June 12-15, Stockholm.

278. Halbert J, Crotty M, Weller D, Ahern M and Silagy C. Primary care-based physical activity programs: effectiveness in sedentary older patients with osteoarthritis symptoms. Arthritis Care and Research. 2001;45:228-234.

279. Halbert J, Weller D, Ahern M and Crotty M. Effectiveness of exercise advice on patients with OA symptoms in a general practice setting. Australian Rheumatology Association. 41st Annual Scientific Conference, Sydney 24-27, May 1998. Scientific Programme. p .73.

280. Hassan BS, Mockett S and Doherty M. Influence of elastic bandage on knee pain, proprioception, and postural sway in subjects with knee osteoarthritis. Annals of the Rheumatic Diseases. 2002;61:24-28.

281. Hassan BS, Mockett S and Doherty M. Static postural sway, proprioception, and maximal voluntary quadriceps contraction in patients with knee osteoarthritis and normal control subjects. Annals of the Rheumatic Diseases. 2001;60(Suppl I):612-618.

282. Häuselmann HJ. Nutripharmaceuticals for osteoarthritis. Best Practice & Research Clinical Rheumatology. 2001;15(4):595-607.

283. Houpt JB, McMillan R, Paget-Dellio D, Russell A and Gahunia HK. Effect of glucosamine hydrochloride (GHcl) in the treatment of pain of osteoarthritis of the knee. The Journal of Rheumatology. 1998;25(Suppl 52):p.8.

284. Houpt JB, McMillan R, Wein C and Paget-Dellio SD. Effect of glucosamine hydrochloride in the treatment of pain of osteoarthritis of the knee. The Journal of Rheumatology. 1999;26:2423-2430.

285. Hughes R and Carr A. A randomized, double-blind, placebo-controlled trial of glucosamine sulphate as an analgesic in osteoarthritis of the knee. Rheumatology. 2002;41:279-284.

286. Hughes RA and Carr AJ. A randomised, double-blind, placebo-controlled trial of glucosamine to control pain in osteoarthritis of the knee. Arthritis and Rheumatism. 2000;43(9)(Suppl):S384.

287. Hughes RA and Carr AJ. A randomised, double-blind, placebo-controlled trial of glucosamine as an analgesic in osteoarthritis of the knee. Rheumatology. 2000;39 (Abstracts Suppl. 1):p.90.

288. Hunsche E, Geling O, Kong SX and Watson DJ. Daily activities of osteoarthritis patients on etoricoxib versus naproxen. Osteoarthritis and Cartilage. 2002;10(Suppl A):S67-S68.

289. Hunsche E, Geling O, Watson DJ and Kong SX. Improvement of daily activities in patients with osteoarthritis treated with etoricoxib, naproxen and placebo. Annals of the Rheumatic Diseases. 2002;61(Supplement). EULAR, June 12-15, Stockholm.

290. Inês LS, Reis P, Santos MJ, Alexandre M, Silva C, Braña A, Barcelos A, Nour D, Silva JAP, Malcata A and Porto A. Joint lavage and infiltration as treatment for knee arthritis: an open follow-up evaluation in the clinical setting. Annals of the Rheumatic Diseases. 2002; 61 (Supplement). EULAR, June 12-15, Stockholm.

291. Jaffe M, Iacobellis D, Young JP, Spiegel K, Martin S, LaMoreaux L, Glessner C, Hess M and Poole RM. Post-hoc results show beneficial effects of pregabalin in patients with osteoarthritis of the hip. Arthritis and Rheumatism. 2000;43(9)(Suppl):S337.

292. Johnson SS, Chaisson CE and Felson DT. Quadriceps strength in knee OA: Relation to painful activities. Arthritis and Rheumatism. 1999;42(9)(Suppl):S145.

293. Johnston LB and Gross MT. The effect of three shoe wear conditions on the lateral patellar displacement in subjects with anterior knee pain. Journal of Orthopaedic & Sports Physical Therapy. 2001;31(1):A3.

294. Kahan A, Guemas E and Lieu PL. Patients with knee osteoarthritis treated by Hylan G-F 20 versus usual treatment: a medico-economic, prospective, randomised large scale trial in France, efficacy and safety outcomes. Annals of the Rheumatic Diseases. 2001;60(Suppl I):232.

295. Kahan A, Lieu P-L and Safin L. Étude prospective randomisée comparant le bénéfice médico-économique de Synvisc à celui dès traitements usuels chez des patients souffrant de gonarthrose. Ruvue du rhumatisme. 2003;70:588-594.

296. Kahan A, Lieu P-L and Safin L. Prospective randomized study comparing the medicoeconomic benefits of hylan GF-20 vs. conventional treatment in knee osteoarthritis. Joint Bone Spine. 2003;70:276-281.

297. Kapusy T, Sunderland M, Lavoie F, Clarke A, Joseph L, Penrod J and De Carolis E. Reduced incidence of adverse events and better dyspepsia-related health during treatment with celecoxib compared to traditional NSAIDs in subjects with osteoarthritis of the hip and/or knee when used in a naturalistic setting. Arthritis and Rheumatism. 2003;48(9)(Suppl):S71.

298. Karim Z, Brown AK, O'Connor PJ, Quinn M, Jackson G, Wakefield RJ, Conaghan PG and Emery P. Do ultrasound-guided corticosteroid injections provide short-term relief of symptoms in patients waiting total hip replacement? Annals of the Rheumatic Diseases. 2001;60(Suppl I):338.

299. Karlsson J and Selin-Sjögren L. A comparison of two hyaluronan drugs and placebo with mild to moderate osteoarthritis of the knee - a controlled, randomised, parallel-design multicenter study. Acta Orthop Scand. 1999;70(Suppl 287):62.

300. Kirkley A, Webster-Bogaert S, Litchfield R, Amendola A, Fowler PJ, McCalden R and MacDonald S. The effect of bracing on medial compartment osteoarthritis of the knee. The Journal of Bone and Joint Surgery. 1999;81-B, suppl.1, p.110.

301. Kirkley A, Webster-Bogaert S, Litchfield R, Amendola A, MacDonald S, McCalden R and Fowler P. The effect of bracing in varus gonarthrosis of the knee. Journal of Bone and Joint Surgery [Am]. 1999;81-A(4):539-548.

302. Kirkley A, Webster-Bogaert S, Litchfield RB, MacDonald S, Amendola A, McCalden R and Fowler PJ. The effect of bracing on medial compartment osteoarthritis of the knee. 66th Annual Meeting Proceedings of the American Academy of Orthopaedic Surgeons. February 4-8, 1999, Anaheim, CA, p.51.

303. Kivitz A, Bensen W, Makarowski W, Bevirt T, Riley W and Recker D. Valdecoxib provides similar efficacy and improved upper GI safety compared with naproxen in elderly and non-elderly patients with osteoarthritis (OA) or rheumatoid arthritis (RA). Arthritis and Rheumatism. 2002;46(9)(Suppl):S160.

304. Kivitz A, Elsen G, Zhao WW, Bevirt T and Recker DP. Randomized placebo-controlled trial comparing efficacy and safety of valdecoxib with naproxen in patients with osteoarthritis. The Journal of Family Practice. 2002;51(6):530-537.

305. Kivitz A, Ma C and Ahdieh H. Oxymorphone extended-release (ER) improves pain and quality of life in patients with osteoarthritis: results of a randomized, double-blind, placebo-controlled, dose-ranging study. EULAR, Lisbon, Portugal, June 18-21, 2003.

306. Klestov A, Bellamy N and Helm H. A double blind, placebo controlled, study of efficacy of action potential stimulation therapy in relieving pain and dysfunction in patients with osteoarthritis of knees. Australian Rheumatology Association, 41st Annual Scientific Conference, Sydney 24-27, May 1998. Scientific Programme. p.116.

307. Klocke R, Jenkinson T, Glew D and Blake DR. Intra-articular ultrasound-guided steroid injection for osteoarthritis of the hip. EULAR 2004, June 9-12, Berlin, Germany.

308. Kocaman O, Koyuncu H, Dinc A and Toros H. Efficiency of electrical stimulation and exercise on knee osteoarthritis. Osteoarthritis and Cartilage. 2003;11(Suppl A):S81.

309. Kotevoglu N, Iyibozkurt P, Hiz Ö and Kuran B. Viscosupplementation treatment for osteoarthritis of the knee. Annals of the Rheumatic Diseases. 2002; 61 (Supplement). EULAR, June 12-15, Stockholm.

310. Koyuncu H, Dinç A, Özkul Y and Yücel E. Clinical, functional and radiological efficacy and tolerability of hyaluronate in osteoarthritis – an open study. Annals of the Rheumatic Diseases. 2002; 61 (Supplement). EULAR, June 12-15, Stockholm.

311. Koyuncu H, Uludag M, Cakmak B, Yucel E, Dinc A and Toros H. The efficacy and safety of viscosupplementation with hylan G-F 20 in osteoarthritic knee. Osteoarthritis and Cartilage. 2003;11(Suppl A):S45.

312. Kriegel W, Korff KJ, Ehrlich JC, Lehnhardt K, Macciocchi A, Moresino C and Pawlowki C. Double-blind study comparing the long-term efficacy of the COX-2 inhibitor nimesulide and naproxen in patients with osteoarthritis. International Journal of Clinical Practice. 2001;55(8):510-514.

313. Krohn K, Yates A, Walczyk J, Buttles S and Berkowitz J. Valgus bracing for isolated medial compartment osteoarthritis of the knee. Arthritis and Rheumatism. 1998.41(9)(Suppl):S198.

314. Kullich WC, Niksic F and Klein G. Effect of nimesulide on metalloproteinases and matrix degradation in osteoarthritis: a pilot clinical study. International Journal of Clinical Practice. 2002. Supplement, Issue 128, July:24-29.

315. Kumar N, Rao C, Marshall NJ and Griffiths ID. Results of one year follow up after treatment with viscosupplementation. EULAR, Lisbon, Portugal, June 18-21, 2003.

316. Kumbhare DA, Conquer J, Barban J, Tannis A, Brown L and Adachi J. Effects of glucosamine on pain and functioning in Canadians with osteoarthritis of the knee. Archives of Physical Medicine and Rehabilitation. 2003;84(9):E10-E11.

317. Kuryata O, Zak MY and Lysunets TK. Celecoxib in the treatment of osteoarthritis in the patients with arterial hypertension: effectiveness. EULAR, Lisbon, Portugal, June 18-21, 2003

318. Kylync S, Gur A, Sarac AJ, Nas K and Cevik R. Comparison of intra-articular hyaluronic acid and physiotherapy in the management of serious knee osteoarthritis. Annals of the Rheumatic Diseases. 2002; 61 (Supplement). EULAR, June 12-15, Stockholm.

319. Lane N, Wolfe F, Sibley J, Luggen M, Oehlert J, Lingala B and Fries J. A longitudinal observational study of progression of knee and hip osteoarthritis (OA): preliminary results on the relationship to NSAID use. Arthritis and Rheumatism. 2002;46(9)(Suppl):S159.

320. Lane N, Wolfe F, Silbey J, Luggen M, Oehlert J, Lingala B and Fries J. Does non-steroidal anti-inflammatory drug (Nsaid) treatment accelerate the progression of osteoarthritis (OA): preliminary results of a randomized control trial of NSAIDS and acetaminophen (APAP). Arthritis and Rheumatism. 2002;46(9)(Suppl):S213.

321. Lane NE, Yan C, Hotary LH, McLaughlin-Miley C and Uprichard MJ. Placebo- and naproxen-controlled dose-ranging trial of darbufelone in osteoarthritis (OA). Arthritis and Rheumatism. 1999;42(9)(Suppl):S404.

322. Lanzer WL and Schuster R. A randomized, single-blind comparison of the efficacy and safety of Synvisc hylan G-F 20 and Aristospan triamcinolone hexacetonide in patients with osteoarthritis (OA) of the knee. American Academy of Orthopaedic Surgeons Annual Meeting, Feb. 13-17, 2002, Dallas.

323. Larsen L, ni15713 Study Group and Stevens RM. Safety of Ro 113-0830, a selective matrix metalloproetease inhibitor (MMPI): a randomized, placebo-controlled trial in patients with knee osteoarthritis (OA). Arthritis and Rheumatism. 2002;46(9)(Suppl):S156.

324. Laurenzi M, Geba GP, Polis AB, Dixon ME, Brady P, Green JA, Littlejohn TW, Dumortier T, Dobbins TW and Gormley GJ. Evaluation of onset of action of rofecoxib, naproxen and nabumetone in patients with osteoarthritis (OA). Annals of the Rheumatic Diseases. 2000;59(Suppl 1).

325. Law A. Diversified chiropractic management in the treatment of osteoarthritis of the knee: a case report. Journal of the Canadian Chiropractic Association. 2001;45(4):232-240.

326. Lee E-O, Song R and Bae S-C. Effects of 12-week tai chi exercise on pain, balance, muscle strength, and physical functioning in older patients with osteoarthritis: randomized trial. Arthritis and Rheumatism. 2001;44(9)(Suppl):S393.

327. Lee E-O, Song R and Bae S-C. Effects of a Sun-style tai chi exercise on motivation and the performance of health behaviours in older women with osteoarthritis. Arthritis and Rheumatism. 2003;48(9)(Suppl):S689.

328. Leeb BF, Bucsi L, Keszthelyi B, Boehmova J, Valesova M, Hawel R and Mayrhofer F. Acemotacin and colocoxib are equally effective and tolerated in the short term treatment of osteoarthritis of the knee joint. Arthritis and Rheumatism. 2003;48(9)(Suppl):S73.

329. Lefkowith JL, Wendt HL, Burr AM, Zhao WW and Geis GS. A comparative study of efficacy and safety in patients with osteoarthritis of the knee receiving either celecoxib, a COX-2 specific inhibitor, or diclofenac. Annals of the Rheumatic Diseases. 2000;59(Suppl 1).

330. Legre V, Pham T and Lafforgue. Evaluation of safety and efficacy of hylan GF-20 in knee osteoarthritis with chondrocalcinosis. Osteoarthritis and Cartilage. 2001;9(Suppl B):S47.

331. Leopold SS, Redd BB, Warme WJ, Wehrle PA, Pettis PD and Shott S. Corticosteroid compared with hyaluronic acid injections for the treatment of osteoarthritis of the knee. A prospective, randomized trial. The Journal of Bone and Joint Surgery. 2003;85-A(7):1197-1203.

332. *Lineker SC, Badley EM, Hawker G and Wilkins A. Determining sensitivity to change in outcome measures used to evaluate hydrotherapy exercise programs for people with rheumatic diseases. Arthritis Care and Research. 2000;13(1):62-65.*

333. *Ma HL, Lin GT, Chang CP, Lac C, Hunsche E and Wang TM. Evaluation of rofecoxib in patients with arthritis in clinical practice in Taiwan – the PEARL study. EULAR 2004, June 9-12, Berlin, Germany.*

334. *Maillefert JF, Hudry C, Baron G, Kieffert P, Bourgeois P, Lechavalier D, Coutaux A and Dougados M. Laterally elevated wedged insoles in the treatment of medial knee osteoarthritis: a prospective randomized controlled study. Osteoarthritis and Cartilage. 2001; 9:738-745.*

335. *Makarowski W, Bevirt T and Recker D. Valdecoxib effectively relieves severe symptoms in patients with OA of the hip. Arthritis and Rheumatism. 2002;46(9)(Suppl):S155.*

336. *Makarowski W, Zhao WW, Bevirt T and Recker DP. Efficacy and safety of the COX-2 specific inhibitor valdecoxib in the management of osteoarthritis of the hip: a randomized, double-blind, placebo-controlled comparison with naproxen. Osteoarthritis and Cartilage. 2002;10(4):290-296.*

337. *Manglone KK, McCully K, Hofmann M and Craik R. Effects of high intensity and low intensity cycle ergometry in older adults with knee osteoarthritis. Arthritis and Rheumatism. 1997;40(9)(Suppl):S86.*

338. *Mangru J, Kumar S and Peloso PM. Is acetaminophen truly safe and effective in OA of the hip and knee? A best evidence synthesis. Arthritis and Rheumatism. 2003;48(9)(Suppl):S80-S81.*

339. *Marcus DM and Suarez-Almazor ME. Is there a role for ginger in the treatment of osteoarthritis? Arthritis and Rheumatism. 2001;44(11):2461-2462.*

340. *Markenson J, Croft J, McCroskery E, Mullins U and Haskell L. Effectiveness of controlled-release oxycontin in osteoarthritis patients with uncontrolled pain moasured by the WOMAC scale. Arthritis and Rheumatism. 2000;43(9)(Suppl):S337.*

341. *Markoll R. Pulsed signal therapy for the treatment of osteoarthritis: double blind and prospective study results in more than 35,000 patients. Osteoarthritis and Cartilage. 1999;7(Supplement A):p.S33.*

342. *Markoll R. Pulsed signal therapy for the treatment of osteoarthritis: double blind and randomized study results in over 50,000 patients. Annals of the Rheumatic Diseases. 2000;59(Suppl).*

343. *Marre J, De Lécluse J, Concas V and Ravaud P. ARGOS: Survey on the association between participation in different types of leisure-sports and articular function in knee femorotibial osteoarthritis. EULAR, Lisbon, Portugal, June 18-21, 2003.*

344. *Martin L, Edworthy SM, Nutting A, MacIntosh B, Butterwick D and Emes C. Exercise in the treatment of osteoarthritis in the knee. Arthritis and Rheumatism. 1995;38(9)(Suppl):S241.*

345. *Mateus M, Pacheco M, Matos AA and Branco JC. Quadriceps isokynetic peak torque in women with knee osteoarthritis. Osteoarthritis and Cartilage. 2000;8(Suppl B):S79-80.*

346. *Matheson AJ and Figgitt DP. Rofecoxib - a review of its use in the management of osteoarthritis, acute pain and rheumatoid arthritis. Drugs. 2001;60(6):833-865.*

347. *Matsumoto A, Babul N, Ahdieh H and Ma T. A placebo- and active-controlled trial of oxymorphone extended release for moderate-to-severe pain in osteoarthritis of the knee or hip. Arthritis and Rheumatism. 2003;48(9)(Suppl):S75.*

348. *Matsumoto A, Ma T, Babul N, Ahdieh H and Lee D. Oxymorphone extended release (ER) improves pain, stiffness, and function in patients with osteoarthritis of the hip or knee. EULAR, Lisbon, Portugal, June 18-21, 2003.*

349. *Matz W, Gilden Tsai A, Garfinkel M, Schumacher HR and Kolasinski SL. The feasibility of using yoga for treating symptoms of osteoarthritis of the knees. Osteoarthritis and Cartilage. 2002;10(Suppl A):S62.*

350. *Mazieres B, Bacri A, Bard H, Le Pen C and Riviere M. French pharmacoeconomic study of hyaluronic acid for osteoarthritis of the knee. An interim analysis. EULAR 2004, June 9-12, Berlin.*

351. *Mazzuca SA, Page MC, Meldrum RD, Brandt KD and Petty-Saphon S. Effects of a heat-retaining sleeve on pain, stiffness, and function in knee osteoarthritis (OA). Arthritis and Rheumatism. 2003;48(9)(Suppl):S423.*

352. *McAlindon TE, Formica MK, Kabbara K, LaValley M and Lehmer M. Results of a protoype internet-based randomized clinical trial (iRCT) of glucosamine for knee osteoarthritis (KOA). Arthritis and Rheumatism. 2003;48(9)(Suppl):S76-S77.*

353. *McKenna F, Borenstein D, Wendt H, Wallemark C, Lefkowith JB and Geis GS. Celecoxib versus diclofenac in the management of osteoarthritis of the knee. Scandinavian Journal of Rheumatology. 2001;30:11-18.*

354. *McKenna F, Weaver A, Fiechtner JJ, Bello AE and Fort JG. COX-2 specific inhibitors in the management of osteoarthritis of the knee: A placebo-controlled, randomized, double-blind study. Journal of Clinical Rheumatology. 2001;7(3):151-159.*

355. *Megyeri A and Genti G. Intra-articular treatment with sodium hyaluronate in osteoarthritis of the knee: Short-term clinical study. XVIIIth Ilar Congress of Rheumatology. Revista Española de Reumatologia. 1993;20(Suppl. 1):329.*

356. *Michalsen A, Klotz S, Lüdtke R, Moebus S, Spahn G and Dobos GJ. Effectiveness of leech therapy in osteoarthritis of the knee. A randomized, controlled trial. Annals of Internal Medicine. 2003;139(9):724-730.*

357. Michalsen A, Klotz S, Spahn G, Lüdtke R, Moebus S and Dobos GJ. Leeches therapy compared to topical NSAID (diclofenac-gel) in the treatment of osteoarthritis of the knee: a randomised controlled trial. EULAR, Lisbon, Portugal, June 18-21, 2003.

358. Michel BA, Stucki G, Haeuselmann H, Frey D, De Vathaire F, Vignon E, Bruehlmann P and Uebelhart D. Retardation of knee joint space narrowing through chondroitin 4 & 6 sulphate. Arthritis and Rheumatism. 2003;48(9)(Suppl):S77.

359. Monfort J and Benito P. Treatment with hyaluronic acid in patients affected with gonarthrosis with and without chondrocalcinosis. Osteoarthritis and Cartilage. 2000; 8(Suppl B):S70.

360. Mongin G, Yakusevich V, Köpe A, Shostak N, Pikhlak E, Popdan L, Simon J, Navarro C, Fortier L, Robertson S and Bouchard S. A safe and efficacious once-daily tramadol formulation for the treatment of pain associated with osteoarthritis of the knee. Osteoarthritis and Cartilage. 2003;11(Suppl A):S117.

361. Moskowitz RW and Cawkwell G. A double-blind, placebo-controlled, randomized 2-week study of valdecoxib 10 mg QD and rofecoxib 25 mg QD in relieving the signs and symptoms of osteoarthritis of the knee in flare. Arthritis and Rheumatism. 2003;48(9)(Suppl):S76.

362. Moskowitz RW, Jones JB and Cawkwell G. Valdecoxib 10 mg demonstrates a rapid onset of action following a single dose in patients with OA of the knee in a flare state. EULAR, Lisbon, Portugal, June 18-21, 2003.

363. Myllykangas-Luosujärvi R, Lu HS, Chen SL, Choon D, Amante C, Chow CT, Pasero G, Genti Gy, Sarembock B, Zerbini CAF, Vrijens F, Moan A, Rodgers DB, De Tora L and Laurenzi M for the Naproxen 901 OF and 901 OC study groups. Comparison of low-dose rofecoxib versus 1000 mg naproxen in patients with osteoarthritis. Scandinavian Journal of Rheumatology. 2002;31:337-344.

364. Nevitt MC, Felson DT, Williams EN, Grady D for the Heart and Estrogen/Progestin Replacement (HERS) Research Group. The effect of estrogen+progestin on knee symptoms (SX) and related disability in postmenopausal women: results from a randomized trial. Arthritis and Rheumatism. 1999;42(9)(Suppl):S146.

365. Nevitt MC, Felson DT, Williams EN, Grady D for the Heart and Estrogen/Progestin Replacement Study Research Group. The effect of estrogen plus progestin on knee symptoms and related disability in postmenopausal women. The Heart and Estrogen/Progestin Replacement Study, a randomized, double-blind, placebo-controlled trial. Arthritis and Rheumatism. 2001;44(4):811-818.

366. O'Reilly S, Muir K and Doherty M. The effect of quadriceps strengthening exercises on pain and disability from osteoarthritis (OA) of the knee. Arthritis and Rheumatism. 1997;40(9)(Suppl):S173.

367. O'Reilly SC, Jones A, Muir KR and Doherty M. Quadriceps weakness in knee osteoarthritis: the effect on pain and disability. Annals of the Rheumatic Diseases. 1998;57:588-594.

368. O'Reilly SC, Muir KR and Doherty M. Effectiveness of home exercise on pain and disability from osteoarthritis of the knee: a randomised controlled trial. Annals of the Rheumatic Diseases.1999;58:15-19.

369. O'Reilly SC, Ratcliffe K, Jones AC, Muir KR and Doherty M. A simple exercise programme can improve pain in knee osteoarthritis: a community based randomised controlled trial. Arthritis and Rheumatism. 1999;42(9)(Suppl):S404.

370. Olszynski WP, Raynauld JP, Goldsmith CH, Torrance GW, Polisson R, Bellamy N, Perlcak D and Walker V. Effectiveness and safety of hylan G-F 20 compared to steroid injections in patients with knee osteoarthritis. Osteoarthritis and Cartilage. 2003;11(Suppl A):S120.

371. Oreizi-Esfahani M, Berman BM, Lao L, et al. Efficacy of traditional Chinese acupuncture in the treatment of osteoarthritis of the knee. Arthritis and Rheumatism. 1995;38(9)(Suppl):S241.

372. Page MC, Mazzuca SA, Brandt KD, Petty-Saphon S and Meldrum RD. Effects of a heat-retaining sleeve on pain, stiffness and function in knee osteoarthritis (OA). American Academy of Orthopaedic Surgeons 71st Annual Meeting, March 10-14, 2004, San Francisco, Poster P196.

373. Paradowski PT, _o_y_ski and Paradowski MT. Clinical follow-up of patients with osteoarthritis of the knee treated with intraarticular pentoxifylline application. Osteoarthritis and Cartilage. 2000;8(Suppl B):S82.

374. Paradowski PT, Morawski K, Szmigiel A and Zolynski K. Randomized double-blind comparison study of rofecoxib and diclofenac in patients with osteoarthritis of the knee. Osteoarthritis and Cartilage. 2001;9(Suppl B):S50.

375. Pavelka K, Bias P, Buchner A, Lammerich A and Schulz U. Licofelone – a 5-LOX, COX-1 and COX-2 inhibitor – is as effective as celecoxib and shows better tolerability in patients with osteoarthritis of the knee. Arthritis and Rheumatism. 2003;48(9)(Suppl):S73.

376. Pavelka K, Bias P, Buchner A, Lammerich A and Schulz U. Licofelone, a 5-lipoxygenase, cyclooxygenase-1 and -2 inhibitor, is as effective as celecoxib and shows improved tolerability in osteoarthritis of the knee. Osteoarthritis and Cartilage. 2003;11(Suppl A):S40.

377. Pavelka K, Bias P, Buchner A, Lammerich A and Schulz U. Licofelone, an inhibitor of COX-1, COX-2 and 5-LOX, is as effective as celecoxib and shows improved tolerability during 12 weeks of treatment in patients with osteoarthritis of the knee. EULAR, Lisbon, Portugal, June 18-21, 2003.

378. Pavelka K, Gatterova J, Olejarova M, Machacek S, Giacovelli G and Rovati LC. Glucosamine sulfate decreases progression of knee osteoarthritis in a long-term, randomized, placebo-controlled, independent, confirmatory trial. Arthritis and Rheumatism. 2000;43(9)(Suppl):S384.

379. Pavelka K, Gatterova J, Olejarova M, Machacek S, Giacovelli G and Rovati LC. A long-term, randomised, placebo-controlled, confirmatory trial on the effects of glucosamine sulfate on knee osteoarthritis progression. Osteoarthritis and Cartilage. 2000;8(Suppl B):S6-7.

380. Pavelka K, Gatterova J, Olejarova M, Machacek S, Gonzalez C, Giacovelli G and Rovati LC. Glucosamine sulfate as an osteoarthritis disease modification agent: a confirmatory, long-term, randomized, placebo-controlled, independent study. Annals of the Rheumatic Diseases. 2001;60(Suppl I):57-58.

381. Pavelka K, Sainerová A, Singh G, Wauk L and Altman R. Collagen intra-articular implant for pain reduction in osteoarthritis of the knee. Annals of the Rheumatic Diseases Abstracts 1999. XIV EULAR Congress, Glasgow, June 6-11, 1999, p.259.

382. Pavelka K, Sainerova A, Singh G, Wauk L and Altman R. Safety following intra-articular injections of Neuvisc - two studies. Osteoarthritis and Cartilage. 1999;7(Supplement A):p.S35.

383. Pelletier J-P, Yaron M, Cohen P for the Investigators in Canada and Israel. Treatment of osteoarthritis of the knee with diacerein: a double-blind, placebo controlled trial. Arthritis and Rheumatism. 1999;42(9)(Suppl):S295.

384. Pelletier J-P, Yaron M, Haraoui B, Cohen P, Nahir MA, Choquette D, Wigler I, Rosner IA, Beaulieu AD, and The Diacerein Study Group. Efficacy and safety of diacerein in osteoarthritis of the knee. A double-blind, placebo-controlled trial. Arthritis and Rheumatism. 2000;43(10):2339-2348.

385. Peloso P, Bellamy N, Bensen W, Thompson G and Babul N. A double blind randomized placebo controlled trial of controlled release codeine in osteoarthritis of hip and knee. Arthritis and Rheumatism. 1997;40(9)(Suppl):S86.

386. Peloso P, Bellamy N, Bensen W, Thompson G and Babul N. A double blind randomized placebo controlled trial of controlled release codeine in osteoarthritis of hip and knee. The Journal of Rheumatology.1998;25(Suppl 52):p.4.

387. Peloso PM, Bellamy N, Bensen W, Thomson GTD, Harsanyi Z, Babul N and Darke AC. Double blind randomized placebo control trial of controlled release codeine in the treatment of osteoarthritis of the hip or knee. The Journal of Rheumatology. 2000;27(3):764-771.

388. Petrella RJ and Bartha C. Home based exercise therapy for older patients with knee osteoarthritis: a randomized clinical trial. The Journal of Rheumatology. 2000;27(9):2215-2221.

389. Petrella RJ, DiSilvestro MD and Cogliano A. Intraarticular hyaluronic acid improves symptoms in "golfer's toe". Osteoarthritis and Cartilage. 2002;10(Suppl A):S71.

390. Petrella RJ, DiSilvestro MD and Hildebrand C. Effects of hyaluronate sodium on pain and physical functioning in osteoarthritis of the knee. A randomized, double-blind, placebo-controlled clinical trial. Archives of Internal Medicine. 2002; 162:292-298.

391. Petrella RJ, DiSilvestro MD and Hildebrand C. Randomized, double-blind, placebo-controlled clinical trial of sodium hyaluronate (Suplasyn) versus NSAID on pain and physical functioning in osteoarthritis of the knee. Annals of the Rheumatic Diseases. 2000;59(Suppl 1).

392. Petronijevic VD, Lazovic MP, Kocic MD and Kojovic ZM. Low-energy laser and pulsed electromagnetic field in the treatment of knee osteoarthritis. Annals of the Rheumatic Diseases. 2002; 61 (Supplement). EULAR, June 12-15, Stockholm.

393. Pham T, Maillefert J, Hudry C, Kieffert P, Bourgeois P, Lechevalier D and Dougados M. Laterally elevated wedged insoles in the treatment of medial knee osteoarthritis. A two-years prospective randomized controlled study. Annals of the Rheumatic Diseases. 2002; 61 (Supplement). EULAR, June 12-15, Stockholm.

394. Pham T, Maillefert J-F, Hudry C, Kieffert P, Bourgeois P, Lechevalier D and Dougados M. Laterally wedged insoles in the treatment of medial knee osteoarthritis. A two-year prospective randomized controlled study. Osteoarthritis and Cartilage. 2004;12:46-55.

395. Pincus T and Sokka T. A pooled index to assess severity of osteoarthritis of the hip or knee. Performance in a randomized, double blind, multicentor, crossover clinical trial of diclofenac/misoprostol (Arthrotec) comparod to acetaminophen in patients with osteoarthritis of the hip or knee. Annals of the Rheumatic Diseases. 2001;60(Suppl I):57.

396. Pincus T and Sokka T. Analysis of worsening of osteoarthritis pain by ≥20% in a randomized controlled double-blind crossover clinical trial of diclofenac/misoprostol (Arthrotec) compared to acetaminophen (ACTA). Arthritis and Rheumatism. 2000;43(9)(Suppl):S339.

397. Pincus T, Callahan LF, Fort J, Drower E, Wilson A, Morant S, Zhao S, Ostorhaus J, Racys V, Lefkowith J and the ACTA study investigators. Comparison of WOMAC scores in osteoarthritis patients in the ACTA [Arthrotec (Diclofenac/Misoprostotol) compared to acetominophen] clinical trial. Arthritis and Rheumatism. 1999;42(9)(Suppl):S74.

398. Pincus T, Callahan LF, Wolfe F, et al. Arthrotec compared to acetaminophen (ACTA): a clinical trial in patients with osteoarthritis (OA) of the hip or knee. Arthritis and Rheumatism. 1999;42(9)(Suppl):S404.

399. Pincus T, Fort JF, Mangal B, Koch G, Wolfe F, Moskowitz R, Simon LS, Gibofsky A, Lei H, Zlotnick S and Sokka T. Patient preference for placebo, acetaminophen (paracetamol), or celecoxib effectiveness study (PACES-1): a double-blind,

randomized, cross-over clinical trial in patients with osteoarthritis of the hip or knee. EULAR, Lisbon, Portugal, June 18-21, 2003.

400. *Pincus T, Gibofsky A, Williams G and McKenna F. Celecoxib 200 mg q.d. and rofecoxib 25 mg qd show similar efficacy in the treatment of a flare of osteoarthritis (OA) of the knee. Annals of the Rheumatic Diseases. 2002; 61 (Supplement). EULAR, June 12-15, Stockholm.*

401. *Pincus T, Koch G, Lei H, Mangal B, Gibofsky A, Moskowitz RW, Simon LS, Wolfe F, Zlotnick S and Fort JG. Patient preference for placebo, acetaminophen or celecoxib: results of 2 crossover, randomized, controlled trials in patients with osteoarthritis of the hip and knee (the PACES Trials). Arthritis and Rheumatism. 2003;48(9)(Suppl):S74-S75.*

402. *Pincus T, Koch GG, Sokka T, Callahan LF, Fort J, Wolfe F, Jordan JM, Luta G, Wang X, Schwartz T, Caldwell JR, Schnitzer TJ, Weaver A and Lefkowith J. Diclofenac/misoprostol (Arthrotec) compared to acetaminophen (ACTA): a clinical trial in patients with osteoarthritis (OA) of the hip or knee. Annals of the Rheumatic Diseases. 2000;59(Suppl 1).*

403. *Pincus T, Koch GG, Sokka T, Lefkowith J, Wolfe F, Jordan JM, Luta G, Callahan LF, Wang X, Schwartz T, Abramson SB, Caldwell JR, Harrell RA, Kremer JM, Lautzenheiser RL, Markenson JA, Schnitzer TJ, Weaver A, Cummins P, Wilson A, Morant S and Fort J. A randomized, double-blind, crossover clinical trial of diclofenac plus misoprostol versus acetaminophen in patients with osteoarthritis of the hip or knee. Arthritis and Rheumatism. 2001;44(7):1587-1598.*

404. *Pincus T, Sokka T and Wolfe F. A visual analog pain scale on a multi-dimensional health assessment questionnaire (MDHAQ) is as sensitive as the WOMAC to document efficacy of diclofenac/misoprostol (Arthrotec) compared to acetaminophen (ACETA) in a randomized controlled crossover clinical trial. Arthritis and Rheumatism. 2001;44(9)(Suppl):S136.*

405. *Pincus T, Sokka T, Koch GG and Wang X. Patient preference in a crossover clinical trial of patients with osteoarthritis of the knee or hip: correlational validity of self-report questionnaire ratings. EULAR 2004, June 9-12, Berlin, Germany.*

406. *Pincus T, Sokka T, Koch GG and Wang X. Validity of a patient preference query on a self-report questionnaire in a randomized controlled crossover clinical trial in patients with osteoarthritis of the hip or knee: consistent results with Western Ontario McMaster Osteoarthritis (WOMAC) scores to indicate greater efficacy of diclofenac/misoprostol versus acetaminophen. Arthritis and Rheumatism. 2003;48(9)(Suppl):S294.*

407. *Pincus T, Sokka T, Wang X and Koch GG. Western Ontario-McMaster Osteoarthritis Index (WOMAC) general versus target joint specific scores indicate similar greater efficacy of diclofenac/misoprostol (Arthrotec) compared to acetaminophen (ACETA) in a randomized controlled crossover clinical trial (ACTA trial). Arthritis and Rheumatism. 2001;44(9)(Suppl):S134.*

408. *Pincus T, Wang X, Sokka T and Koch GG. Patients with greater severity of osteoarthritis of the hip or knee according to Western Ontario McMaster Osteoarthritis Scale (WOMAC) scores have greater differences between responses to diclofenac/misoprostol (Arthrotec) versus acetaminophen on all WOMAC subscales in a randomized crossover clinical trial. Arthritis and Rheumatism. 2001;44(9)(Suppl):S134.*

409. *Pipitone N and Scott DL. Magnetic pulse treatment for knee osteoarthritis: a randomised, double-blind, placebo-controlled study. Current Medical Research and Opinion. 2001;17(3):190-196.*

410. *Power JD, Badley E, Cott C and Hawker G. Determinants of referral to physical therapy services for people with at least moderately severe hip or knee arthritis. Arthritis and Rheumatism. 2002;46(9)(Supplement):S405.*

411. *Quilty B, Tucker M and Dieppe PA. A randomized controlled trial of physiotherapy for patellofemoral joint osteoarthritis of the knee (PFJ OA). Osteoarthritis and Cartilage. 2000;8(Suppl B):S67.*

412. *Quilty B, Tucker M and Dieppe PA. Physiotherapy for patellofemoral joint osteoarthritis of the knee. A randomised controlled trial. Arthritis and Rheumatism. 1998;41(9)(Suppl):S199.*

413. *Quilty B, Tucker M, and Dieppe PA. The relationship between pain, disability and quadriceps strength in patello-femoral joint osteoarthritis. Annals of the Rheumatic Diseases Abstracts 1999. XIV EULAR Congress, Glasgow, June 6-11, 1999, p.256.*

414. *Quilty B, Tucker M, Campbell R and Dieppe P. Physiotherapy, including quadriceps exercises and patellar taping, for knee osteoarthritis with predominant patello-femoral joint involvement: randomized controlled trial. The Journal of Rheumatology. 2003;30(6):1311-1317.*

415. *Rainsford KD, Bolten W, Schuchlein K, Dempsey A and Schnitker J. Effects of intramuscular sodium ribonucleianate [Osteochondrin S] in osteoarthritis of the knee: identification of "responders" and "non-responders". EULAR 2004, June 9-12, Berlin, Germany.*

416. *Rao C, Kumar N, Young-Min S, Marshall NJ, Platt PN and Griffiths ID. Experiences with a viscosupplementation (Hyalgan) clinic. Annals of the Rheumatic Diseases. 2001;60(Suppl I):233.*

417. *Ravaud P, Giraudeau B, Logeart I, Larguier JS, Treves R, Euller-Ziegler L, Bannwarth B and Dougados M. Management of osteoarthritis (OA) with patient-administered assessment tools and/or unsupervised home-based exercises program. A 2x2 factorial design cluster randomized controlled trial. Arthritis and Rheumatism. 2001;44(9)(Suppl):S355.*

418. *Ravaud P, Giraudeau B, Logeart I, Larguier JS, Treves R, Euller-Ziegler L, Bannwarth B and Dougados M. Management of osteoarthritis (OA) with patient-administered assessment tools and/or unsupervised home-based exercises program. A 2x2 factorial design cluster randomized controlled trial. Osteoarthritis and Cartilage. 2001;9(Suppl B):S47-48.*

419. Raynauld JP, Bellamy N, Goldsmith C, Tugwell P, Torrance G, Walker V, Schultz M, Pericak D and Band P. A prospective randomized health outcomes trial of hylan G-F 20 in the treatment of patients with knee osteoarthritis. Arthritis and Rheumatism. 2000;43(9)(Suppl):S340.

420. Raynauld JP, Bellamy N, Goldsmith CH, Tugwell P, Torrance GW, Pericak D, Band P, Polisson R and Bolovich D. An evaluation of the safety and effectiveness of repeat courses of hylan G-F 20 for treating patients with knee osteoarthritis. Osteoarthritis and Cartilage. 2002;10(Suppl A):S63.

421. Raynauld J-P, Buckland-Wright C, Tremblay J-L, Khy V, Bertrand C, Choquette D, Haraoui B, Martel-Pelletier J and Pelletier J-P. Clinical trials: impact of intra-articular steroid injections on the progression of knee osteoarthritis. Osteoarthritis and Cartilage. 2000;8(Suppl B):S16-17.

422. Raynauld J-P, Buckland-Wright C, Ward R, Choquette D, Haraoui D, Martel-Pelletier J, Uthman I, Khy V, Tremblay J-L, Bertrand C and Pelletier J-P. Safety and efficacy of long-term intraarticular steroid injections of the knee. A randomized, double-blind, placebo-controlled trial. Arthritis and Rheumatism. 2003;48(2):370-377.

423. Raynauld JP, Olszynski WP, Goldsmith CH, Torrance GW, Bellamy N, Polisson R, Pericak D and Walker V. Effectiveness and safety of hylan G-F 20 compared to steroid injections in patients with knee osteoarthritis. Arthritis and Rheumatism. 2003;48(9)(Suppl):S484.

424. Raynauld JP, Torrance GW, Band PA, Goldsmith CH, Tugwell P, Walker V, Schultz M, Bellamy N. A prospective, randomized, pragmatic, health outcomes trial evaluating the incorporation of hylan G-F 20 into the treatment paradigm for patients with knee osteoarthritis (Part 1 of 2): clinical results. Osteoarthritis and Cartilage 2002;10(7):506-17.

425. Raynauld J-P, Torrance GW, Bellamy N, Goldsmith CH, Tugwell P, Walker V, Schultz M and Band P. The effectiveness, cost-effectiveness and cost-utility of Hylan G-F 20 in osteoarthritis of the knee: clinical results. The Journal of Rheumatology. 2000;27(7):1814.

426. Raynauld J-P. Clinical trials: impact of intraarticular steroid injections on the progression of knee osteoarthritis. Osteoarthritis and Cartilage. 1999;7:348-349.

427. Redd BB, Leopold SS, Warme WJ, Pettis PD, Wehrle PA and Shott S. Corticosteroid vs. Synvisc injections for knee arthritis: a randomized, controlled trial demonstrating similar overall efficacy but significant gender-related treatment differences. American Academy of Orthopaedics Surgeons 70th Annual Meeting, Feb. 5-9, 2003, New Orleans.

428. Reginster J, Bias P and Buchner A. First clinical results of licofelone (ML3000), an inhibitor of COX-1, COX-2 and 5-LOX, for the treatment of osteoarthritis. Annals of the Rheumatic Diseases. 2002; 61 (Supplement). EULAR, June 12-15, Stockholm.

429. Reginster J, Rovati L, Deroisy R, Henrotin Y, Gatterova J, Machacek S, Gonzalez C, Giacovelli G and Pavelka K. Glucosamine sulfate slows down osteoarthritis progression in postmenopausal women: pooled analysis of two large, independent, randomized, placebo-controlled, double-blind, prospective 3-year trials. Annals of the Rheumatic Diseases. 2002; 61 (Supplement). EULAR, June 12-15, Stockholm.

430. Reginster J-Y, Deroisy R, Lee RL, Rovati LC, Lejeune E, Giacovelli G, Henrotin Y, Dacre J and Gosset C. Long-term disease modifying effect by glucosamine sulfate in knee osteoarthritis: relationships between structure and symptom changes. Annals of the Rheumatic Diseases. 2000;59(Suppl 1).

431. Reginster JY, Deroisy R, Paul I, Lee RL, Giacovelli G, Dacre J, Rovati LC and Gosset C. Glucosamine sulfate significantly reduces progression of knee osteoarthritis over 3 years: a large randomized, placebo-controlled, double-blind, prospective trial. Osteoarthritis and Cartilage. 2000;8(2):151-152.

432. Reginster JY, Deroisy R, Paul I, Lee RL, Henrotin Y, Giacovelli G, Dacre J, Rovati LC and Gosset C. Long-term disease modifying effect by glucosamine sulfate in knee osteoarthritis: relationships between structure and symptom changes. Rheumatology. 2000;39(Abstracts Suppl. 1):p.91.

433. Reginster JY, Deroisy R, Rovati LC, Lee RL, Lejeune E, Bruyere O, Giacovelli G, Henrotin Y, Dacre JE and Gossett C. Long-term effects of glucosamine sulfate on osteoarthritis progression: a randomized, placebo-controlled clinical trial. Lancet 2001;Jan.27;357(9252):251-6. Comment in Lancet 2001;Jan.27;357(9252):247-248; reviewed in: The Journal of Family Practice, May 2001;50(5).

434. Reginster JY, Fischer CL, Beualieu A, Malmstrom K, Curtis SP, Miller ER, Bergman G, Smolar J and Reicin AS. Etoricoxib demonstrates similar efficacy and improved gastrointestinal safety compared to naproxen in two 138-week randomized studies of osteoarthritis patients. EULAR 2004, June 9-12, Berlin, Germany.

435. Renkl[inodot]tepe N and Atalay E. The effect of intra-articular sodium hyaluronate therapy in knee osteoarthritis. Annals of the Rheumatic Diseases. 2000;59(Suppl 1).

436. Roddy JE, Carroll GJ, Surveyor I, Will RK, Bell MC and Wang H. 90Yttrium does not suppress refractory synovitis in rheumatoid arthritis (RA). Arthritis and Rheumatism. 1998;41(Suppl.9):p.S315.

437. Rosemffet MG, Schneeberger EE, Citera G, Gagliardi S and Maldonado-Cocco JA. Effects of functional electrostimulation on pain, muscular strength and functional capacity in patients with osteoarthritis of the knee. Arthritis and Rheumatism. 2003;48(9)(Suppl):S641.

438. Roth SH and Shainhouse JZ. A U.S., multicenter, placebo-controlled, double-blinded, clinical trial of diclofenac topical solution (Pennsaid®) in the treatment of the symptoms of primary osteoarthritis of the knee. Annals of the Rheumatic Diseases. 2002; 61 (Supplement). EULAR, June 12-15, Stockholm.

439. *Russell ID, Baker D and Johnson SR. A comparison of Synvisc and arthroscopic lavage in the management of osteoarthritis of the knee. The Journal of Bone and Joint Surgery [Br]. 2003;85-B(Suppl II):106.*

440. *Saag K, Fisher C, McKay J, et al. MK-0966, a specific COX-2 inhibitor, has clinical efficacy comparable to ibuprofen in the treatment of knee and hip osteoarthritis (OA) in a 6-week controlled clinical trial. Arthritis and Rheumatism. 1998;41(9)(Suppl):S196.*

441. *Sangdee C, Teekachunhatean S, Sananpanich K, Sugandharesa N, Chiewchantanakit K, Pojchamarnwiputh S and Jayasvasti S. Electroacupuncture versus diclofenac in symptomatic treatment of osteoarthritis of the knee: a randomized controlled trial. BMC Complementary and Alternative Medicine 2002;2:3.*

442. *Schell E, Boucher L, Tamasi L, Ouellet J, Moore A and Lee A. Long-term efficacy and tolerability of lumiracoxib in osteoarthritis of the knee. EULAR, Lisbon, Portugal, June 18-21, 2003.*

443. *Schmid B, Lüdtke R, Selbmann H-K, Kötter I, Tschirdewahn B, Schaffner W et al. Efficacy and tolerability of a standardized willow bark extract in patients with osteoarthritis: randomized, placebo-controlled, double blind clinical trial. Wirksamkeit und Verträglichkeit eines standardisierten Weidenrindenextraktes bei Arthrose-Patienten: Randomisierte, placebo-kontrollierte doppelblindstudie. Zeitschrift für Rheumatologie. 2000;59(5):314-320.*

444. *Schmid B, Lüdtke R, Selbmann H-K, Kötter I, Tschirdewahn B, Schaffner W and Heide L. Efficacy and tolerability of a standardized willow bark extract in patients with osteoarthritis: randomized placebo-controlled, double blind clinical trial. Phytotherapy Research. 2001;15(4):344-350.*

445. *Schnitzer TJ, Geusens P, Hasler P, Patel SK, Poor G, Senftleber I, Jayawardene S, Sloan VS and Macerata RS. Efficacy and safety of COX 189 in osteoarthritis: a multinational study. Arthritis and Rheumatism. 2000;43(9)(Suppl):S336.*

446. *Schnitzer TJ, Kivitz AJ, Greenwald M, Fleischmann RM, Matzura-Wolfe D, Polis AB, Dixon ME, Dobbins TW and Geba GP. Rofecoxib provides superior relief of symptoms of osteoarthritis (OA) compared to celecoxib. Annals of the Rheumatic Diseases. 2001;60(Suppl I):236.*

447. *Schnitzer TJ, Kivitz AJ, Lipetz RS, Sanders N and Hee A. A phase II study of the efficacy and safety of AZD3582, a CINOD, in subjects with osteoarthritis of the knee. EULAR 2004, June 9-12, Berlin, Germany.*

448. *Sepici V, Sepici A, Acikgoz S and Tali T. The role of viscosupplementation with sodium hyaluronate (Orthovisc) in urine GAG levels and magnetic resonance imaging (MRI) features of osteoarthritis of the knee. Annals of the Rheumatic Diseases. 2002; 61 (Supplement). EULAR, June 12-15, Stockholm.*

449. *Shariati J, Hatef M and Jokar M. Intra-articular hylan G-F 20 (Synvisc) in the treatment of patients with osteoarthritis of the knee. The Journal of Rheumatology. 2001;28(Suppl 63):3 (ILAR, Edmonton, 2001).*

450. *Shin Y, Park K, Kim S, Do J, Cho C and Kim W. Comparison of efficacy between non-steroidal anti-inflammatory drug and acetaminophen in patients with knee osteoarthritis, based on ultrasonographic severity. EULAR 2004, June 9-12, Berlin, Germany.*

451. *Shuba NM, Kovalenko VM, Voronova TD, Bortkevich OP and Slavskaya NV. Efficacy of treatment of knee osteoarthritis by preparation diacerein within 12 months. EULAR, Lisbon, Portugal, June 18-21, 2003.*

452. *Siollwończyk P, Mazurkiewicz S, Ostojski R and Kusiak A. Ocena wartości dostawowych iniekcji preparatu Hyalgan w leczeniu zmian zwyrodnieniowych stawu kolanowego [The value of intraarticular hyalgan injections in treatment of osteoarthritis of the knee]. Chir Narz Ruchu Ortop Pol 1997;LXIII(4):337-342.*

453. *Silverfield JC, Kamin M, Wu S-C, Rosenthal N for the CAPSS-105 Study Group. Tramadol/Acetaminophen combination tablets for the treatment of osteoarthritis flare pain: a multicenter, outpatient, randomized, double-blind, placebo-controlled, parallel-group, add-on study. Clinical Therapeutics. 2002;24(2):282-297.*

454. *Singer F, Mayrhofer F, Klein G, Hawel R and Kollenz CJ. Evaluation of the efficacy and dose-response relationship of dexibuprofen (S(+)-ibuprofen) in patients with osteoarthritis of the hip and comparison with racemic ibuprofen using the WOMAC osteoarthritis index. International Journal of Clinical Pharmacology and Therapeutics. 2000;38(1):15-24.*

455. *Singh G and Fort J. Efficacy of celecoxib and rofecoxib: is there any clinically important difference? Annals of the Rheumatic Diseases. 2002; 61 (Supplement). EULAR, June 12-15, Stockholm.*

456. *Singh G, Fort JG, Triadafilotoulos G and Bello A. SUCCESS-1: a global osteoarthritis (OA) trial in 13,274 randomized patients celecoxib provides similar efficacy to diclofenac and naproxen while providing significantly improved UGI safety. Arthritis and Rheumatism. 2001;44(9)(Suppl):S135.*

457. *Singh G, Goldstein J, Bensen W, Agrawal N, Elsen G, Fort J, Bellow A and Boots S. SUCCESS-1 in osteoarthritis (OA) trial: celecoxib significantly reduces risk of serious upper GI complications compared to NSAIDs while providing similar efficacy in 13,274 randomized patients. Annals of the Rheumatic Diseases. 2001;60(Suppl I):58.*

458. *Singh GK, Jain S, Das SK and Sharma V. Conservative treatment of osteoarthritis (OA) of the knee vs add-on intra-articular (IA) steroid injection. Journal of Clinical Epidemiology. 1996;49(1)(Suppl 1):p.8S.*

459. *Singh GS, Gagnier P, Hwang E, Krishnan E and Kobe M. A treatment strategy based on consistent use of meloxicam 7.5 mg achieves better efficacy compared to a "usual clinical care" strategy in patients with osteoarthritis: results from the IMPROVE trial, a multi-center, randomized parallel-group, open label study of 1309 patients in a managed care setting. Annals of the Rheumatic Diseases. 2002; 61 (Supplement). EULAR, June 12-15, Stockholm.*

460. *Sivri A. Quality of life in osteoarthritis patients treated with nonsteroidal anti-inflammatory drugs versus analgesics. Scandinavian Journal of Rheumatology. 2000;Suppl. 114, p.40.*

461. *Sokka T and Pincus T. A visual analog pain scale on a Multi-Dimensional Health Assessment Questionnaire (MDHAQ) is as sensitive as the WOMAC to document efficacy of diclofenac/misoprostol (Arthrotec) compared to acetaminophen in a randomized controlled crossover clinical trial. Annals of the Rheumatic Diseases. 2002; 61 (Supplement). EULAR, June 12-15, Stockholm.*

462. *Sokka T and Pincus T. Greater efficacy for diclofenac/misoprostol (Arthrotec) compared to acetaminophen in a randomized, double blind, multicenter, crossover clinical trial in patients with osteoarthritis of the hip or knee is most evident in patients with the most severe osteoarthritis. Annals of the Rheumatic Diseases. 2001;60(Suppl I):228.*

463. *Sokka T and Pincus T. Greater efficacy of diclofenac/misoprostol (Arthrotec) compared to acetaminophen in a randomized, double blind, crossover clinical trial is most evident in patients with more severe osteoarthritis (OA) of the hip or knee. The Journal of Rheumatology. 2001;28(Suppl 63):4 (ILAR, Edmonton, 2001).*

464. *Sokka T, Pincus T, Fort J and Lefkowith J. Analyses of 3 measures of patient responses indicating greater efficacy of diclofenac/misoprostol (Arthrotec) versus acetaminophen in a randomized, controlled, crossover osteoarthritis clinical trial. Annals of the Rheumatic Diseases. 2000;59(Suppl 1).*

465. *Sokka T, Pincus T, Wang X and Koch GG. The 'target joint' in patients with osteoarthritis of the hip or knee was not identical in 17% of patients in 5 visits over 14 weeks in randomized crossover controlled clinical trial to compare diclofenac/misoprostol versus acetaminophen. Arthritis and Rheumatism. 2003;48(9)(Suppl):S296.*

466. *Song R, Lee E-O, Lam P and Bae S-C. Effects of tai chi exercise on pain, balance, muscle strength, and perceived difficulties in physical functioning in older women with osteoarthritis: a randomized clinical trial. The Journal of Rheumatology. 2003;30(9):2039-2044.*

467. *Soroka NF and Chyzh KA. Clinical efficiency and pharmacoeconomical evaluation of treatment by chondroitin sulphate (Structum®) in patients with primary osteoarthritis (POA). Annals of the Rheumatic Diseases. 2002; 61 (Supplement). EULAR, June 12-15, Stockholm.*

468. *Sowers J, White W, Pitt B, Whelton A, Simon L, Van Ingen H and Fort J. Rofecoxib, but not celecoxib, at doses providing equal OA efficacy, increases mean 24-hr systolic blood pressure in treated hypertensive patients with osteoarthritis and type 2 diabetes mellitus. EULAR, Lisbon, Portugal, June 18-21, 2003.*

469. *Spector TD, Conaghan P, Buckland-Wright JC, Cline GC, Beary JF and Meyer JM. Risedronate produces disease modification and symptomatic benefit in the treatment of knee osteoarthritis: results from the BRISK study. Arthritis and Rheumatism. 2003;48(12):3650-3651.*

470. *Sprott H, Shen H, Michel BA, Uebelhart D, Gay RE, Gay S and Sprott H. Analgesic action of acetaminophen and rofecoxib on knee pain and function. EULAR, Lisbon, Portugal, June 18-21, 2003.*

471. *Stengaard-Pedersen K, Ekesbo R, Karvonen A-L and Lyster M. Celecoxib 200 mg q.d. is efficacious in the management of osteoarthritis of the knee or hip regardless of the time of dosing. Rheumatology. 2004;43(5):592-595.*

472. *Takeda W and Wessel J. Acupuncture for the treatment of pain of osteoarthritic knees. Arthritis Care and Research. 1994;7:118-122.*

473. *Tamir E, Robinson D, Koren R, Agar G and Halperin N. Intra-articular hyaluronan injections for the treatment of osteoarthritis of the knee: a randomized, double-blind, placebo controlled study. Clinical and Experimental Rheumatology. 2001;19:265-270.*

474. *Tannenbaum H, Berenbaum F, Reginster J-Y, Zacher J, Robinson J, Poor G, Bliddal H, Uebelhart D, Adadmi S, Navarro F, Lee A, Moore A and Gimona A. Lumiracoxib (Prexige) is effective in the treatment of osteoarthritis of the knee: a 13-week, placebo-controlled, active-comparator, double-blind study. Arthritis and Rheumatism. 2002;46(9)(Suppl):S212.*

475. *Tekeoglu I, Adak B, Goksoy T and Tosun N. Effects of intra-articular injections of sodium hyaluronate (Orthovisc) and betamethasone on osteoarthritis of the knee. The Journal of Rheumatology & Medical Rehabilitation. 1998;9(4):220-224.*

476. *Tekeoglu I, Ediz L and Adak B. Comparison the effects of acetaminophen and colchicine in the management of primary knee osteoarthritis. EULAR 2004, June 9-12, Berlin, Germany.*

477. *Thamsborg GM, Apel K, Rein E, Kharazmi A and Winther K. A randomized, double-blind, placebo-controlled, cross-over trial of the herbal remedy hybenvital in patients with osteoarthritis. Annals of the Rheumatic Diseases. 2002; 61 (Supplement). EULAR, June 12-15, Stockholm.*

478. *Theiler R, Bischoff H, Good M and Uebelhart D. Rofecoxib improves quality of life in hip and knee OA patients. Osteoarthritis and Cartilage. 2002;10(Suppl A):S68-S69.*

479. *Theiler R, Uebelhart D, Bischoff H and Good M. Rofecoxib improves quality of life in hip and knee OA patients. EULAR, Lisbon, Portugal, June 18-21, 2003.*

480. *Thomas KS, Muir KR, Doherty M, Jones AC, O'Reilly SC, Bassey EJ on behalf of the Community Osteoarthritis Research Group. Home based exercise programme for knee pain and knee osteoarthritis: randomized controlled trial. BMJ. 2002;325:752-755. (Abridged version; full version on bmj.com)*

481. *Thompson JI, Huang YW and Zaibel R. Safety and efficacy of fermentation-derived high molecular weight sodium hyaluronate – a clinical trial in patients with osteoarthritis of the knee. Osteoarthritis and Cartilage. 2002;10(Suppl A):S70-S71.*

482. *Toh EM, Prasad P and Teanby D. Correlating the outcome of knee intraarticular viscosupplementation with radiological changes of osteoarthritis. The Journal of Bone and Joint Surgery [Br]. 2003;85-B(Suppl II):178.*

483. *Toh EM, Prasad PS and Teanby D. Correlating the efficacy of knee viscosupplementation with osteoarthritic changes on roentgenological examination. The Knee. 2002;9:321-330.*

484. *Torrance GW, Bellamy N, Raynauld JP, Goldsmith CH, Polisson R, Belovich D, Pericak D and Gough J. A prospective, randomized controlled trial evaluating the effectiveness and safety of hylan G-F 20 in patients with knee osteoarthritis: six month clinical results of a single course of treatment. Osteoarthritis and Cartilage. 2002;10(Suppl A):S63.*

485. *Torrance GW, Raynauld J-P, Bellamy N, Goldsmith CH, Tugwell P, Walker V, Schultz M and Band P. The effectiveness, cost-effectiveness and cost-utility of Hylan G-F 20 in osteoarthritis of the knee: economic results. The Journal of Rheumatology. 2000;27(7):1814.*

486. *Torrance GW, Raynauld JP, Walker V, Goldsmith CH, Bellamy N, Band PA, Schultz M, Tugwell P. A prospective, randomized, pragmatic, health outcomes trial evaluating the incorporation of hylan G-F 20 into the treatment paradigm for patiens with knee osteoarthritis (Part 2 of 2): economic results. Osteoarthritis and Cartilage 2002;10(7):518-27.*

487. *Truitt K, Ettinger W Jr., Schnitzer T, Greenwald M, Daniels B, Zeng Q, Seidenberg B and Ehrich E. Rofecoxib, a cox-2 specific inhibitor, had clinical efficacy and overall safety in treating osteoarthritis patients aged 80 years and older. Annals of the Rheumatic Diseases Abstracts 1999. XIV EULAR Congress, Glasgow, June 6-11, 1999, p.206.*

488. *Tsai C-L, Chang C-C, Chen S-C, Beinat L and Piva S. Treatment of knee osteoarthritis in Asian population with an intra-articular hyaluronan of MW 500-730 KDA. Osteoarthritis and Cartilage. 2003;11(Suppl A):S119.*

489. *Tsoukas C, Eyster E, Mukhopadhyay S, Giallella K and Melian A. Efficacy of etoricoxib in the treatment of hemophilic arthropathy. Arthritis and Rheumatism. 2001;44(9)(Suppl):S124.*

490. *Tugwell PS, Wells GA and Shainhouse JZ. A comparative trial of oral versus topical diclofenac in the treatment of osteoarthritis of the knee. EULAR, Lisbon, Portugal, June 18-21, 2003.*

491. *Tukmachi E, Dempsey E and Jubb RW. Effect of acupuncture in knee osteoarthritis. Osteoarthritis and Cartilage. 2000;8(Suppl B):S75.*

492. *Uebelhart D, Knüssel O and Theiler R. Efficacy and tolerability of oral avian chondroitin sulfate in painful knee osteoarthritis: a double-blind, placebo-controlled, multicentre 6-month trial. Osteoarthritis and Cartilage. 1999;7(Supplement A):p.S36.*

493. *Waddell DD, Cefalu CA and Bricker DC. A 52-week, open-label study of a second course of hylan G-F 20 for the treatment of pain associated with osteoarthritis of the knee. EULAR, Lisbon, Portugal, June 18-21, 2003.*

494. *Waddell DD, Cefalu CA and Bricker DC. An open-label study of a second course of hylan G-F 20 for the treatment of pain associated with knee osteoarthritis. Current Medical Research and Opinion. 2003;19(6):499-507.*

495. *Weaver AL, Polis AB, Petruschke RA, Najarian DK and Geba GP. Onset of efficacy of rofecoxib compared to nabumetone and placebo in patients with osteoarthritis. EULAR, Lisbon, Portugal, June 18-21, 2003.*

496. *Weaver AL, Polis AB, Petruschke RA, Najarian DK and Geba GP. Time to onset of efficacy of rofecoxib compared to nabumetone or placebo in patients with osteoarthritis. Osteoarthritis and Cartilage. 2003;11(Suppl A):S81.*

497. *Weaver AL, Schnitzer TJ, Polis AB, Petruschke RA and Geba GP. Efficacy and safety of rofecoxib compared to celecoxib or acetaminophen among 1960 osteoarthritis patients: the VACT trials. Osteoarthritis and Cartilage. 2003;11(Suppl A):S42.*

498. *Weaver AL, Schnitzer TJ, Polis AB, Petruschke RA, Keane WF and Geba GP. Efficacy and safety of rofecoxib compared to celecoxib or acetaminophen among 1960 osteoarthritis patients: the VACT trials. Arthritis and Rheumatism. 2003;48(9)(Suppl):S75.*

499. *Weaver AL. Rofecoxib: clinical pharmacology and clinical experience. Clinical Therapeutics. 2001;23(9):1323-1338.*

500. *Wein CR, Houpt JB, McMillan R and Gahunia HK. Open trial of glucosamine hydrochloride (GHcl) in the treatment of pain of osteoarthritis of the knee. The Journal of Rheumatology. 1998;25(Suppl 52): p.8.*

501. *Weiss C. Pain management with hylan G-F 20 in patients awaiting total knee arthroplasty. Source Unknown.*

502. *Wigler MD, Grotto I, Caspi D and Yaron M. The effects of Zintona EC (a ginger extract) on symptomatic gonarthritis. Osteoarthritis and Cartilage. 2003;11:783-789.*

503. *Williams GW, Ettlinger RE, Ruderman EM, Hubbard RC, Lonien ME, Yu SS, Zhao W and Geis GS. Treatment of osteoarthritis with a once-daily dosing regimen of celecoxib. A randomized, controlled trial. Journal of Clinical Rheumatology. 2000;6(2):65-74.*

504. *Williams GW, Hubbard RC, Yu SS, Zhao W and Geis GS. Comparison of once-daily and twice-daily administration of celecoxib for the treatment of osteoarthritis of the knee. Clinical Therapeutics. 2001;23(2):213-227.*

505. *Wislowska M. Efficiency and safety of selective cyclooxygenase-2 (COX-2) inhibition (Meloxicam) therapy in comparison with diclofenac in osteoarthritis. Osteoarthritis and Cartilage. 2000;8(Suppl B):S87.*

506. *Wittenberg RH, Schell E, Krehan G, Maeumbacd R, Runge H, Schlüter P, Fashola TO, Thurston HJ, Burger K, Trechsel U and Gimona A. Prospective, randomized study on the first-dose analgesic effect of lumiracoxib in osteoarthritis of the knee. EULAR, Lisbon, Portugal, June 18-21, 2003.*

507. *Wluka AE, Stuckey S, Brand C and Cicuttini FM. Supplementary vitamin E does not affect the loss of cartilage volume in knee osteoarthritis: a 2 year double blind randomized placebo controlled study. The Journal of Rheumatology. 2002;29(12):2585-2591.*

508. *Wolfe F, Flowers N, Burke T, Arguelles L and Pettitt D. Increase in lifetime adverse drug reactions and increased disease severity among patients who will start COX-2 agents: quantitative assessment of confounding by indication in 6,689 RA and OA patients. Arthritis and Rheumatism. 2000;43(9)(Suppl):S225.*

509. *Wolfe F, Flowers N, Burke TA, Arguelles LM and Pettitt D. Increase in lifetime adverse drug reactions, service utilization, and disease severity among patients who will start COX-2 specific inhibitors: quantitative assessment of channeling bias and confounding by indication in 6689 patients with rheumatoid arthritis and osteoarthritis. The Journal of Rheumatology. 2002;29:1015-1022.*

510. *Yazici H, Ozdemir O, Kirazli Y, Ozbek S, Ozerbil O, Gur A et al. Celecoxib (200 mg/d) or diclofenac (100 mg/d) in osteoarthritis: a multicentre, double-blind trial. EULAR, Lisbon, Portugal, June 18-21, 2003.*

511. *Yocum D, Fleischmann R, Dalgin P, Caldwell J, Hall D, Roszko P for the Meloxicam Osteoarthritis Investigators. Safety and efficacy of meloxicam in the treatment of osteoarthritis. A 12-week, double-blind, multiple-dose, placebo-controlled trial. Archives of Internal Medicine. 2000;160:2947-2954.*

512. *Yocum DE, Hall DB and Roszko PJ. Safety and efficacy of meloxicam in the treatment of osteoarthritis (OA): results of a 12 week double-blind, placebo controlled trial. Annals of the Rheumatic Diseases. 2000;59(Suppl 1).*

513. *Yocum DE, Hall DB, Roszko PJ and the Meloxicam OA Team. Efficacy and safety of meloxicam in the treatment of osteoarthritis (OA): results of a phase III double-blind, placebo controlled trial. Arthritis and Rheumatism. 1999;42(9)(Suppl):S147.*

514. *Yocum DE, McGoldrick K, Waller PR, Pavesi M, Segarra R and Páez A. LAS 34475, a novel COX-2 inhibitor, shows a superior rate of responders compared to naproxen in osteoarthritis of the knee. EULAR, Lisbon, Portugal, June 18-21, 2003.*

515. *Yocum DE, McGoldrick K, Waller PR, Segarra R, Páez A, Pavesi M and Musoles S. LAS 34475, a novel COX-2 inhibitor, shows high potency at low doses in patients with osteoarthritis of the knee. EULAR, Lisbon, Portugal, June 18-21, 2003.*

516. *Yocum DE, Musoles S, Paez A, Segarra R, McGoldrick K and Waller P. Results of a phase II, double blind, placebo controlled trial with an active comparator using LAS 34475, a highly selective COX-2 inhibitor in osteoarthritis of the knee. Arthritis and Rheumatism. 2003;48(9)(Suppl):S78.*

517. *Young L, Katrib Am Cuello C, Vollmer-Conna U, Bertouch JV, Roberts-Thomson PJ, Ahern MJ, Smith MD and Youssef PP. Effects of intraarticular glucocorticoids on macrophage infiltration and mediators of joint damage in osteoarthritis synovial membranes. Findings in a double-blind, placebo-controlled study. Arthritis and Rheumatism. 2001;44(2):343-350.*

518. *Zacher J and Schattenkirchner M. Osteoarthritis: results of ADAM (Advanced Arthritis Management), a survey in Germany. Annals of the Rheumatics Diseases. 2002; 61(Supplement). EULAR, Stockholm, June 12-15.*

519. *Zacher J, Feldman D, Gerli R, Scott D, Hou S-M, Uebelhart D, Rodger IW and Ozturk ZE for the Etoricoxib OA Study Group. A comparison of the therapeutic efficacy and tolerability of etoricoxib and diclofenac in patients with osteoarthritis. Current Medical Research and Opinion. 2003;19(8):725-736.*

520. *Zenk JL, Helmor TR, Kuskowski MA. The effects of milk protein concentrate on the symptoms of osteoarthritis in adults: an exploratory, randomized, double-blind, placebo-controlled trial. Current Therapeutic Research 2002;63(7):430-42.*

521. *Zhao S, Fiechtner JI, Tindall EA, Dedhiya SD, Zhao WW, Osterhaus JT and Yu SS. Evaluation of health-related quality of life of rheumatoid arthritis patients treated with celecoxib. Arthritis Care and Research. 2000;13(2):112-121.*

522. *Zhao SZ, Dedhiya SD, Verburg K and Osterhaus JT. Celecoxib 200 mg administered once-a-day or in split doses has equal impact on health-related quality of life (HRQOL) of patients with osteoarthritis (OA). Arthritis and Rheumatism. 1999;42(9)(Suppl):S297.*

523. *Zhao SZ, Dedhiya SD, Verburg K, Fort J and Osterhaus JT. Celecoxib improves health related quality of life (HRQOL) of elderly patients with osteoarthritis (OA). Arthritis and Rheumatism. 1999;42(9)(Suppl):p.S296.*

524. *Zhao SZ, McMillen JI, Markenson JA, Dedhiya SD, Zhao WW, Osterhaus JT and Yu SS. Evaluation of the functional status aspects of health-related quality of life patients with osteoarthritis treated with celecoxib. Pharmacotherapy. 1999;19(11):1269-1278.*

525. *Zhao SZ, Yu SS, Dedhiya SD, Zhao WW and Osterhaus JT. Celecoxib improves health-related quality of life in patients with osteoarthritis of the knee. Pharmacotherapy. 1999;19(4):495.*

526. *Zhao ZS, Yu SS, Zhao WW and Osterhaus JT. Celecoxib reduces pain and improves functioning in patients with osteoarthritis (OA). Annals of the Rheumatic Diseases Abstracts 1999, XIV EULAR Congress, Glasgow, June 6-11, 1999, p.204.*

527. Zochling J, March L, Lapsley H, Tribe K, Cross M and Brooks P. Complementary medicines - are our OA patients spending their money wisely? Osteoarthritis and Cartilage. 2000;8(Suppl B):S86-87.

Applications of the WOMAC in Orthopaedic Surgery Studies.

528. Adili A, Bhandari M, Giffin R, Whately C and Kwok DC. Valgus high tibial osteotomy. Comparison between an Ilizarov and a Coventry wedge technique for the treatment of medial compartment osteoarthritis of the knee. Knee Surgery, Sports Traumatology, Arthroscopy. 2002;10(3):169-176.

529. Agnidis Z, Slimec J, Krajbich J, Hutchison CR and Gross AE. Health-related quality-of-life following fresh osteochondral allograft of the knee: a minimum five year follow-up. The Journal of Bone and Joint Surgery. 1999;81-B(Suppl.I):p.106.

530. Ahn H, Manzary M, Koo J and Mahomed N. Prospective study of functional outcomes of arthroscopic surgery for knee osteoarthritis. The Canadian Orthopaedic Association 56th Annual Meeting, London, Canada, June 1-4, 2001. Programme Book, p. 116.

531. Almqvist KF, Verdonk P, Verdonk R, Verstraete K, Verbruggen G and Veys EM. Are biodegradable matrices seeded with mature human chondrocytes the future treatment for articular cartilage defects? EULAR, Lisbon, Portugal, June 18-21, 2003.

532. Al-Sayyad MJ and Cameron JC. Functional outcome after tibial tubercle transfer for the painful patella alta. Clinical Orthopaedics and Related Research. 2002;No.396:152-162.

533. Al-Sobayel HI, Cooke TDV, Harrison L, Khan B and Chaudhry A. Outcome evaluation of arthritic Arabic patients with knee implant. The Canadian Orthopaedic Association 55th Annual Meeting, June 3-6, 2000, Edmonton, Alberta. Programme 2000, p.50.

534. Amin S, LaValley MP, Niu J, Sco GS, Hill CL, Gale, D, Gale ME and Felson DT. Complete anterior cruciate ligament (ACL) tear and risk for radiographic and symptom progression in subjects with knee osteoarthritis. Arthritis and Rheumatism. 2003;48(9)(Suppl):S70-S71.

535. Anderson JG, Wixson RL, Tsai D, Stulberg SD and Chang RW. Functional outcome and patient satisfaction in total knee patients over the age of 75. The Journal of Arthroplasty. 1996;11:831-840.

536. Ang DC, Ibrahim SA, Burant CJ, Siminoff LA and Kwoh CK. Ethnic differences in the perception of prayer and consideration of joint arthroplasty. Medical Care. 2002;40(6):471-476.

537. Arnold WJ. Polyurethane joint restoration treatment of osteoarthritis. 65th Annual Meeting Proceedings of the American Academy of Orthopaedic Surgeons. February 4-8, 1999, Anaheim, CA, p.429-430.

538. Ashraf T, Newman JH, Evans RL and Ackroyd CE. Lateral unicompartmental knee replacement. Survivorship and clinical experience over 21 years. The Journal of Bone and Joint Surgery. 2002;84-B:1126-1130.

539. Ashworth MA. The development of an orthopedic waiting list. Canadian Health Services Research Foundation. March 29, 2004. http:www.chsrf.ca.

540. Ayers DC, Jain R, Rogers M, Franklin PD, Vanglynn M, Bertolo G and Murray D. Patient outcome measures in TKR patients: preoperative mental function affects postoperative physical function. American Academy of Orthopaedic Surgeons Annual Meeting, Feb. 13-17, 2002, Dallas.

541. Ayers DC, Jain RK, Bennett MT, Setter K, Rogers M, Franklin PD and Murray DG. WOMAC scores improve predictive measures for TJA. The Journal of Arthroplasty. 2000;15(2):p.254.

542. Ayers DC, Jain RK, Rogers M, Franklin PD, Van Glynn M, Bertolo G and Murray DG. How patient outcome measures in TKR patients can be useful in your practice. The Journal of Arthroplasty. 2001;16(2):253.

543. Bachmeier CJM, March LM, Cross M, et al. A comparison of outcome in men and women with OA undergoing joint replacement as measured with the WOMAC. Australian Rheumatology Association. 40th Annual Scientific Conference, Brisbane 25-28, May 1997. Scientific Programme. p .156.

544. Bachmeier CJM, March LM, Cross MJ, Lapsley HM, Tribe KL, Courtenay BG, Brooks PM for the Arthritis Cost and Outcome Project Group. A comparison of outcomes in osteoarthritis patients undergoing total hip and knee replacement surgery. Osteoarthritis and Cartilage. 2001;9:137-146.

545. Banwart JC, McQueen DA, Long MJ, Rizk WS and Schurman II JR. The effect of flexion-extension gap symmetry on functional outcome in total knee arthroplasty. American Academy of Orthopaedic Surgeons 71st Annual Meeting, March 10-14, 2004, San Francisco, Poster P151.

546. Barton-Hanson NG, Finley R and Parkinson RW. Poor validity of the two most commonly used outcome questionnaires in total knee replacement. Validity of: Hospital for Special Surgery (HSS), Knee Society (KSS) and Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) scores compared. The Journal of Bone and Joint Surgery. 1999;81-B(Suppl.II):p.216.

547. Beaupre L, Davies DM, Jones CA and Cinats JG. Exercise combined with continuous passive motion or slidor board therapy compared with exercise only: a randomized controlled trial of patients following total knee arthroplasty. Physical Therapy. 2001;81(4):1029-1037.

548. Beaupre LA, Lier D, Davies DM and Johnston DBC. The effect of a preoperative exercise and education program on functional recovery, health related quality of life, and health service utilization following primary total knee arthroplasty. The Journal of Rheumatology. 2004;31(6):1166-1173.

549. Bederman S, Schemitsch EH and Waddell JP. How do traditional hip scoring systems correlate with validated disease-specific and general health outcome measures? The Canadian Orthopaedic Association 55th Annual Meeting, June 3-6, 2000, Edmonton, Alberta. Programme 2000, p.170.

550. Bederman S, Schemitsch EH, Waddell JP, et al. Functional outcome following periprosthetic fractures around total hip arthroplasty. The Canadian Orthopaedic Association 55th Annual Meeting, June 3-6, 2000, Edmonton, Alberta. Programme 2000, p.131.

551. Berven S, Psychoyios V, Crawford RW, Murray DW, Hayes J and Lingard E. A comparison of preoperative disease severity in patients undergoing total joint arthroplasty under two different health care delivery systems. 66th Annual Meeting Proceedings of the American Academy of Orthopaedic Surgeons. February 4-8, 1999, Anaheim, CA, p.8.

552. Bhattacharyya T, Gale D, Dewire P, Trotterman S, Elon Gale M, Einhorn TA and Felson D. The clinical significance of meniscal tears on osteoarthritis. American Academy of Orthopaedic Surgeons Annual Meeting, Feb. 13-17, 2002, Dallas.

553. Bischoff HA, Lingard EA, Losina E, Roos E, Philipps CB, Mohamed NN, Baron JA, Barrett J and Katz JN. Mental health, common geriatric problems, obesity and education are independant cross-sectional correlates of functional outcomes 3 years after total hip replacement. Arthritis and Rheumatism. 2001;44(9)(Suppl):S184.

554. Blanchard C, Feeny D, Mahon JL, Bourne R, Rorabeck C, Stitt L and Webster-Bogaert S. Is the Health Utilities Index responsive in total hip arthroplasty patients? Journal of Clinical Epidemiology. 2003;56:1046-1054.

555. Boardman DL, Dorey F, Thomas BJ and Lieberman JR. The accuracy of assessing total hip arthroplasty outcomes. The Journal of Arthoplasty. 2000;15(2):200-204.

556. Bourne R and Rorabeck CH. Primary femoral fixation options. Assessing the outcomes: what really works. Orthopedics. 1999;22(9):823-825.

557. Bourne RB and Rorabeck CH. A critical look at cementless stems. Taper designs and when to use alternatives. Clinical Orthopaedics and Related Research. 1998;No.355:212-223.

558. Bourne RB, Mahon J, Rorabeck CH, Feeny D, Stitt L and Webster-Bogaert S. The effect of waiting your turn for elective total hip arthroplasty (THA) for osteoarthritis. The Canadian Orthopaedic Association 55th Annual Meeting, June 3-6, 2000, Edmonton, Alberta. Programme 2000, p.166.

559. Bourne RB, Rorabeck CH, Laupacis A, et al. A randomized clinical trial comparing cemented to cementless total hip replacement in 250 osteoarthritic patients: the impact on health related quality of life and cost effectiveness. The Iowa Orthopaedic Journal. 1994;14;108-114.

560. Bowen CVA, Botsford DJ, Hudak PL and Evans PJ. Microsurgical treatment of septic nonunion of the tibia. Quality of life results. Clinical Orthopaedics and Related Research. 1996;No.332:52-61.

561. Braeken AM, Lochhaas-Genach JA, Gollish JD, Myles JM and MacKenzie TA. Determinants of 6-12 month postoperative functional status and pain after elective total hip replacement. International Journal for Quality in Health Care. 1997;9(6):413-418.

562. Bremander ABI, Petersson IF and Roos E. Validation of the Rheumatoid Arthritis Outcome Score (RAOS) – for the lower extremity. EULAR, Lisbon, Portugal, June 18-21, 2003.

563. Bremander ABI, Petersson IF and Roos EM. Validation of the rheumatoid and arthritis outcome score (RAOS) – for the lower extremity. Arthritis and Rheumatism. 2003;48(9)(Suppl):S687.

564. Brion PH, Moreland LW, Myers S, Ike RW, Klashman DJ, Concoff AL, Rich E, Skovron ML and Kalunian KC. Visually-guided irrigation in patients with early knee osteoarthritis: a multicenter randomized, controlled trial. Arthritis and Rheumatism. 2000;43(9)(Suppl):S338.

565. Brokelman RBG, van Loon CJM and Rijnberg WJ. Patient versus surgeon satisfaction after total hip arthroplasty. The Journal of Bone and Joint Surgery. 2003;85-B(4):495-498.

566. Bullens PHJ, van Loon CJM, de Waal Malefijt MC, Laan RFJM and Veth RPH. Patient satisfaction after total knee arthroplasty. A comparison between subjective and objective outcome assessments. The Journal of Arthroplasty. 2001;16(6):740-747.

567. Burnett RS, Haydon C, Elkassem H, Rorabeck CH and Bourne RB. Patellar resurfacing/nonresurfacing in total knee arthroplasty: a randomized trial at minimum 10 yrs. American Academy of Orthopaedic Surgeons 71st Annual Meeting, March 10-14, 2004, San Francisco, Paper Presentation, Paper No. 010.

568. Burnett RS, Yuan X, Yang KY, MacDonald SJ, McCalden RW, Bourne RB, Rorabeck CH and O'Brien JJ. Isolated liner exchange in revision THA: reduced dislocation rate with the direct lateral approach. American Academy of Orthopaedic Surgeons 71st Annual Meeting, March 10-14, 2004, San Francisco, Poster No. P087.

569. Callaghan JJ and Johanson NA. Gender influences on SF-36 and WOMAC scores for total hip and total knee arthroplasty. American Academy of Orthopaedic Surgeons 68th Annual Meeting, San Francisco, U.S.A. (Poster PE146).

570. Case JP, Ryals AR, Schnitzer TJ, Hurwitz DE and Block JA. Validity of a modified Hospital for Special Surgery (HSS) knee score in medically-treated osteoarthritis (OA). Arthritis and Rheumatism. 2000;43(9)(Suppl):S110.

571. Chang HJ, Levy PS and Hughes SL. Analyzing the gender gap in timing of knee and hip replacement surgery. Arthritis and Rheumatism. 2003;48(9)(Suppl):S222-S223.

572. Chen AL, Mujtaba M, Zuckerman JD, Jeong GK, Joseph TN, Wright K and Di Cesare PE. Midterm clinical and radiographic results with the Genesis I total knee prosthesis. The Journal of Arthroplasty. 2001;16(8):1055-1062.

573. Cinats JG, Beaupre LA, Davies DM, Jones CA, Suarez-Almazor ME and Johnston DWC. Comparison of three mobilization methods following total knee arthroplasty (TKA). The Journal of Bone and Joint Surgery [Br]. 2000;82-B(Suppl.I):p.4.

574. Clark CR, Rorabeck CH, MacDonald S and MacDonald DA. Posterior sacrificed versus cruciate retaining implants in TKA: results from a multicenter randomized trial. American Academy of Orthpaedic Surgeons 68th Annual Meeting, Feb. 28 - Mar. 4, 2001, San Francisco, U.S.A. (Paper 044).

575. Clark CR, Rorabeck CH, MacDonald S, MacDonald DA, Swafford J and Cleland D. Posterior-stabilized and cruciate-retaining total knee replacement. Clinical Orthopaedics and Related Research. 2001;No.392:208-212.

576. Clark M, Masson E, O'Connor GJ, Beaupre L, Lavoie GJ and Scharfenberger A. Further experience with the Prostalac hip system at an alternate site. American Academy of Orthopaedic Surgeons 71st Annual Meeting, March 10-14, 2004, San Francisco, Poster No. P012.

577. Concoff A, Moreland L, Skovron ML, et al. Office-based arthroscopic irrigation in early knee osteoarthritis: results of a multi-center study. Arthritis and Rheumatism. 1997;40(9)(Suppl):S86.

578. Cross MJ, March LM, Lapsley HM, Tribe KL, Bachmeier CJ, Courtenay BG and Brooks PM. Determinants of willingness to pay for hip and knee joint replacement surgery for osteoarthritis. Osteoarthritis and Cartilage. 1999;7(Supplement A):p.S26.

579. Cross MJ, March LM, Lapsley HM, Tribe KL, Brnabic AJM, Courtenay BG and Brooks PM. Determinants of willingness to pay for hip and knee joint replacement surgery for osteoarthritis. Rheumatology. 2000;39:1242-1248.

580. Currey SS, Dent JL, Callahan LF, Jonas BI, Nuss G, Esther RJ, Almekinders LC and Garrett WE. Pain and patients' perceived outcome following arthroscopic knee surgery (AKS). Arthritis and Rheumatism. 2001;44(9)(Suppl):S156.

581. Dapuzzo MR, Hernandez VH, Diaz RM, Sierra RJ and Lavernia CJ. Mental health and outcome in primary joint replacement of the hip and knee. American Academy of Orthopaedic Surgeons 71st Annual Meeting, March 10-14, 2004, San Francisco, Poster Number P026.

582. Davey JP and Santore RF. Complications of periacetabular osteotomy. Clinical Orthopaedics and Related Research. 1999;No.363:33-37.

583. Davies AP. Rating systems for total knee replacement. The Knee. 2002;9:261-266.

584. Dawson J and Carr A. Outcomes evaluation in orthopaedics. Journal of Bone and Joint Surgery. 2001;83-B,313-315.

585. de la Serna AR and Castro NB. Joint lavage are beneficial pain treatment in knee osteoarthritis. The Journal of Rheumatology. 2001;28(Suppl 63):7 (ILAR, Edmonton, 2001).

586. Dunbar MJ, Obertsson O and Ryd L. Making soft data hard: results of reliability studies of outcome questionnaires for knee arthroplasty patients. Journal of Bone and Joint Surgery. 2001;83-B, Suppl. I:16-17.

587. Dunbar MJ, Robertson O, Ryd, L and Lidgren L. The knee bone is connected to the hip bone... Journal of Bone and Joint Surgery. 2001;83-B,Suppl. I:16.

588. Dunbar MJ, Robertsson O, Ryd L and Lidgren L. Appropriate questionnaires for knee arthroplasty. Results of a survey of 3600 patients from the Swedish knee arthroplasty registry. Journal of Bone and Joint Surgery. 2001;83-B(3):339-344.

589. Dunbar MJ, Robertsson O, Ryd L and Lidgren L. Making soft data hard: results of reliability studies on outcome questionnaires for knee arthroplasty. The Canadian Orthopaedic Association 55th Annual Meeting, June 3-6, 2000, Edmonton, Alberta. Programme 2000, p.79.

590. Dunbar MJ, Robertsson O, Ryd L and Lidgren L. The knee bone is connected to the hip bone... The Canadian Orthopaedic Association 55th Annual Meeting, June 3-6, 2000, Edmonton, Alberta. Programme 2000, p.76.

591. Dunbar MJ, Robertsson O, Ryd L and Lidgren L. Translation and validation of the Oxford-12 item knee score for use in Sweden. Acta Orthop Scand. 2000;71(3):268-274.

592. Dunbar MJ, Valdivia GG, Bourne RB and Rorabeck CH. Why collect pre-operative WOMAC scores for knee arthroplasty? Journal of Bone and Joint Surgery. 2001;83-B, Suppl. I:16.Dunbar MJ, Valdivia GG, Rorabeck CH and Bourne RB. Why collect preoperative WOMAC scores for knee arthroplasty? The Canadian Orthopaedic Association 55th Annual Meeting, June 3-6, 2000, Edmonton, Alberta. Programme 2000, p.76.

593. Dunbar MJ. Subjective measures after knee arthroplasty. Thesis, Lund, Sweden. Acta Orthopaedica Scandinavica. 2001;Supplementum No. 301, vol. 72:1-63.

594. El Maraghy A, Schemitsch EH, Tsang G, Morton J and Waddell JP. Mid-term functional outcome of uncemented total hip arthroplasty for avascular necrosis. The Canadian Orthopaedic Association 55th Annual Meeting, June 3-6, 2000, Edmonton, Alberta. Programme 2000, p.130.

595. Englund M, Roos EM, Roos HP and Lohmander LS. Patient-relevant outcome fourteen years after meniscectomy: influence of type of meniscal tear and size of resection. Rheumatology. 2001;40:631-639.

596. Ethgen O, Bruyère O, Richy F, Dardennes C and Reginster J-Y. Health-related quality of life in total hip and total knee arthroplasty. The Journal of Bone and Joint Surgery. 2004;86-A(5):963-974.

597. Ethgen O, Tancredi A, Lejeune E, Kvasz A, Zegels B and Reginster J-Y. Do utility values and willingness to pay suitably reflect health outcome in hip and knee osteoarthritis? A comparative analysis with the WOMAC Index. The Journal of Rheumatology. 2003;30(11):2452-2459.

598. Ethgen O, Tancredi A, Lejeune E, Zegels B, Kvazs A and Reginster J-Y. Do willingness to pay and time trade-off approaches suitably reflect health outcomes in hip and knee osteoarthritis? Arthritis and Rheumatism. 2002;42(9)(Suppl):S73.

599. Ethgen O, Tancredi A, Lejeune E, Zegels, Kvazs A and Reginster J-Y. Do willingness to pay and time trade-off approaches suitably reflect health outcomes in hip and knee osteoarthritis? Arthritis and Rheumatism. 2002;46(9)(Suppl):S73.

600. Fielden JM, Cumming JM, Horne JG, Devane PA, Gallagher LM and Slack A. Waiting for hip arthroplasty: economic costs and health outcomes. The Journal of Bone and Joint Surgery [Br]. 2003;85-B(Suppl III):202.

601. Finch E, Walsh M, Thomas SG and Woodhouse LJ. Functional ability perceived by individuals following total knee arthroplasty compared to age-matched individuals without knee disability. Journal of Orthopaedic and Sports Physical Therapy. 1988;27:255-263.

602. Fortin PR, Clarke AE, Joseph L, et al. A transnational comparison of preoperative status and outcomes of elective total hip/knee replacement. Arthritis and Rheumatism. 1996;39(9)(Suppl):S209.

603. Fortin PR, Clarke AE, Joseph L, et al. Outcomes of total hip and knee replacement. Preoperative functional status predicts outcomes at six months after surgery. Arthritis and Rheumatism. 1999;42(8):1722-1728.

604. Fortin PR, Joseph L, Tanzer M, et al. Worse general health and more severe disease in Canadian vs American patients undergoing elective total hip/knee surgery for osteoarthritis. Arthritis and Rheumatism. 1995;38(9)(Suppl):S382.

605. Friedman RJ, Longo III JA, Newman MA, Jessup DE and Bal BS. Prospective randomized study comparing ceramic/ceramic and ceramic/polyethylene bearing surfaces in total hip arthroplasty. American Academy of Orthopaedic Surgeons Annual Meeting, Feb. 13-17, 2002, Dallas.

606. Gangji V, Matos C, Lambermont M, Appelboom T and Hauzeur J-P. Autologous bone marrow transplantation and core decompression in the treatment of osteonecrosis of the femoral head. Arthritis and Rheumatism. 2002;46(9)(Suppl):S227.

607. Garbuz D and Van Bergeyk A. Quality of life and sports-specific outcome after Bernese periacetabular osteotomy. American Academy of Orthopaedic Surgeons Annual Meeting, Feb. 13-17, 2002, Dallas.

608. Garbuz D. Bernese periacetabular osteotomy provides pain relief and increased function for active patients. AAOS, Dallas TX, Feb. 18, 2001.

609. Giammattei FP and Giammattei CP. Palm Pilot data entry system. The Journal of Arthroplasty. 2002;17(2):248. American Association of Hip and Knee Surgeons 11th Annual Meeting, Nov. 9-11, 2001, Dallas, Texas.

610. Goetz DD, Vail TP, Tanzer M, Fisher DA, Callaghan JJ and Mohler CG. A prospective randomized trial of cemented femoral components with polished versus grit blasted surface finish and identical stem geometry. American Academy of Orthopaedics Surgeons 70th Annual Meeting, Feb. 5-9, 2003, New Orleans.

611. Groll DL, Rudan J and MacKenzie TA. A one-year follow-up study of patient-perceived health status following total knee replacement surgery. The Journal of Bone and Joint Surgery. 1999;81-B(Suppl.I):p.106.

612. Haddad FS, Garbuz DS, Chambers GK, Jagpal TJ, Duncan CP and Masri BA. The expectation of patients undergoing revision hip arthroplasty. The Canadian Orthopaedic Association 55th Annual Meeting, June 3-6, 2000, Edmonton, Alberta. Programme 2000, p.132-133.

613. Haddad FS, Garbuz DS, Chambers GK, Jagpal TJ, Masri BA and Duncan CP. The expectations of patients undergoing revision hip arthroplasty. The Journal of Arthroplasty. 2001;16(1):87-91.

614. Haddad FS, Garbuz DS, Chambers GK, Jagpal TJ, Masri BA and Duncan CP. The expectations of patients undergoing revision hip arthroplasty. The Journal of Bone and Joint Surgery [Br]. 2003;85-B(Suppl I):35.

615. Haddad FS, Marko A, McEwen J, Garbuz DS, Masri BA and Duncan CP. Variations between estimated and actual pain-activity relationships in lower limb arthroplasty candidates: implications for improved assessment of patient outcomes. The Journal of Bone and Joint Surgery [Br]. 2000;82-B(Suppl.I):p.73.

616. Haddad FS, Marko A, McEwen J, Garbuz DS, Masri BA and Duncan CP. The pain-activity relationship as an outcome indicator for arthroplasty surgery. The Journal of Bone and Joint Surgery [Br]. 2000;82-B(Suppl.II):p.126.

617. Harrison MM, Morrell J and Hopman W. The influence of obesity on patient satisfaction and function after arthroscopic knee surgery. The Canadian Orthopaedic Association 55th Annual Meeting, June 3-6, 2000, Edmonton, Alberta. Programme 2000, p.103.

618. Hartford JM, Hunt T and Kaufer H. Cemented versus uncemented mobile bearing total knee replacement: results at 5 to 13-year follow-up. 67th Annual Meeting Proceedings, American Academy of Orthopaedic Surgeons, March 15-19, 2000. Orlando, Florida, p.352.

619. Hartley R, Barton-Hanson N, Finley R and Parkinson R. Patient outcomes in primary and revision knee replacement. A prospective study. The Journal of Bone and Joint Surgery [Br]. 2003;85-B(Suppl):15.

620. Hartley RC, Barton-Hanson NG, Finley R, Parkinson RW. Early patient outcomes after primary and revision total knee arthroplasty. The Journal of Bone and Joint Surgery 2002;84-B(7):994-999.

621. Hauser DL, Wessinger SJ, Condon RT, Golladay GJ, Hoeffel DP, Gillis DJ, Merrill DR, Chaisson D, Freiberg AA, Estok DM, Rubash HE, Malchau H and Harris WH. An electronic database for outcome studies that includes digital radiographs. The Journal of Arthroplasty. 2001;16(8):Suppl. 1:71-75.

622. Hawker G, Badley E, Coyte P, et al. Determining the correct rate for joint replacement. Arthritis and Rheumatism. 1998;41(9)(Suppl):S76.

623. Hawker G, Wright J, Coyte P, Badley E, Williams J, Harvey B and Glazier R. Sex differences in arthroplasty provision. Arthritis and Rheumatism. 1999;42(9)(Suppl):S72.

624. Hawker G, Wright J, Coyte P, Paul J, Dittus R, Croxford R, Katz B, Bombardier C, Heck D and Freund D. Health-related quality of life after knee replacement. The Journal of Bone and Joint Surgery. 1998;80-A(2):163-173.

625. Hawker GA, Badley EM, Wright JG and Coyte PC. Understanding willingness to consider hip/knee joint arthroplasty. Arthritis and Rheumatism. 2002;46(9)(Suppl):S73.

626. Hawker GA, Bombardier C, Freund D, et al. U.S. survey of knee replacement (KR) recipients to assess short and long-term outcomes. Arthritis and Rheumatism. 1993;36(9)(Suppl):S85.

627. Hawker GA, Wright JG, Badley EM, Coyte PC and Glazier RH. Education, income and unmet need for total joint arthroplasty. Arthritis and Rheumatism. 2001;44(9)(Suppl):S325.

628. Hawker GA, Wright JG, Coyte PC, Williams JI, Harvey B, Glazier R and Badley EM. Differences between men and women in the rate of use of hip and knee arthroplasty. The New England Journal of Medicine. 2000; 342:1016-1022.

629. Hawker GA, Wright JG, Coyte PC, Williams JI, Harvey B, Glazier R, Wilkins A and Badley EM. Determining the need for hip and knee arthroplasty: the role of clinical severity and patients' preferences. Medical Care. 2001;39(3):206-216.

630. Heck DA, Robinson RL, Partridge CM, Lubitz RM and Freund DA. Patient outcomes after knee replacement. Clinical Orthopaedics and Related Research. 1998;356:93-110.

631. Hewins E, Leighton RK, Petrie DP and the Multicentre Arthroplasty Group. A randomized clinical trial comparing modular versus non-modular total knee replacements. The Canadian Orthopaedic Association 56th Annual Meeting, London, Canada, June 1-4, 2001. Programme Book, p.77.

632. Horton TC, Lingard EA and Moran CG. The effect of psychological factors on outcome following total knee replacement. The Journal of Bone and Joint Surgery [Br]. 2003;85-B(Suppl II):95.

633. Huber E. Assessment of quality of life after total knee arthroplasty. Rheumatology in Europe. 1998;27(Suppl 2):179.

634. Ibrahim SA, Siminoff LA, Burant CJ, Kwoh CK. Differences in expectations of outcome mediate African American/White patient differences in "willingness" to consider joint replacement. Arthritis and Rheumatism 2002;46(9):2429-2435.

635. Institute for Clinical Evaluative Sciences. ICES Technical Report #00-04-TR "Information strategy: urgency rating, waiting list management and patient outcomes monitoring for primary hip/knee joint replacement. August 2000. The Working Group: P. Blackstein-Hirsch, J.D. Gollish, G. Hawker, H. Kreder, N. Mahomed, J.I. Williams for the ICES Consensus Group.

636. Jain R, Koo M, Kreder HJ, Schemitsch EH, Davey JR and Mahomed NN. Avascular necrosis and subcapital hip fractures in patients <60 years. Journal of Bone and Joint Surgery. 2001;83-B, Suppl. I:17.

637. Jain R, Koo M, Kreder HJ, Schemitsch EH, Davey JR and Mahomed NN. Avascular necrosis and subcapital hip fractures in patients less than 60 years. The Canadian Orthpaedic Association 55th Annual Meeting, June 3-6, 2000, Edmonton, Alberta. Programme 2000, p.82-83.

638. Jain R, Koo M, Kreder HJ, Schemitsch EH, Davey JR, Mahomed N and McKee MD. Avascular necrosis and subcapital hip fractures in patients less than 60 years of age. American Academy of Orthopaedic Surgeons 68th Annual Meeting, Feb. 28 - Mar. 4, 2001, San Francisco, U.S.A. (Poster PE048).

639. Jain R, Koo M, Kreder HJ, Schemitsch EH, Davey JR, Mahomed NN. Comparison of early and delayed fixation of subcapital hip fractures in patients sixty years of age or less. The Journal of Bone and Joint Surgery 2002;84-A(9):1605-1612.

640. Jeer PJS, Atrey A, Conry BG and Gibb PA. Patellofemoral pain and function following total knee arthroplasty. The Journal of Bone and Joint Surgery [Br]. 2003;85-B(Suppl I):56.

641. Johanson NA, Liang MH, Daltroy L, Rudicel S and Richmond J. American Academy of Orthopaedic Surgeons lower limb outcomes assessment instruments. Reliability, validity, and sensitivity to change. The Journal of Bone and Joint Surgery. 2004;86-A(5):902-909.

642. Johnston RC and Callaghan JJ. The pros and cons of a computer based patient record. The Journal of Arthroplasty. 2002;17(2):246. American Association of Hip and Knee Surgeons 11th Annual Meeting, Nov.9-11, 2001, Dallas, Texas.

643. Jones A, Voaklander D, Johnston DWC, Kramer G, Redfern L and Suarez-Almazor M. Outcome of hip and knee arthroplasties in an aged population. A population-based cohort study. Arthritis and Rheumatism. 1997;40(9)(Suppl): S327.

644. Jones CA, Voaklander D, Johnston DWC and Suarez-Almazor ME. Determinants of function after total knee arthroplasty. Arthritis and Rheumatism. 1998;41(9)(Suppl):S360.

645. Jones CA, Voaklander D, Johnston DWC and Suarez-Almazor ME. Baseline status and perioperative factors do not predict functional improvement or pain relief in patients undergoing hip and knee arthroplasties. Arthritis and Rheumatism. 1998;41(9)(Suppl):S357.

646. Jones CA, Voaklander D, Johnston DWC and Suarez-Almazor ME. Pain, function, and overall health status after total hip (THA) and knee (TKA) arthroplasties in a population-based cohort. Arthritis and Rheumatism. 1998;41(9)(Suppl):S91.

647. Jones CA, Voaklander D, Johnston DWC and Suarez-Almazor ME. The effect of age on pain, function, and quality of life after total hip and knee arthroplasty. Archives of Internal Medicine. 2001;161:454-460.

648. Jones CA, Voaklander DC, Johnston DWC and Suarez-Almazor ME. Determining factors of patient satisfaction after total hip (THA) and knee arthroplasties (TKA). Arthritis and Rheumatism. 1999;42(9)(Suppl):S268.

649. Jones CA, Voaklander DC, Johnston DWC and Suarez-Almazor ME. Health related quality of life outcomes after total hip and knee arthroplasties in a community based population. The Journal of Rheumatology. 2000;27(7):1745-1752.

650. Jones CA, Voaklander DC, Johnston DWC and Suarez-Almazor ME. The relationship of range of motion and the WOMAC with total hip (THA) and knee arthroplasties (TKA). Arthritis and Rheumatism. 1999;42(9)(Suppl):S286.

651. Kalunian K, Klashman D, Singh R, et al. Office-based arthroscopy in early osteoarthritis (OA) of the knee: preliminary results of a multi-center study of the effects of visually-guided irrigation on outcome. Arthritis and Rheumatism. 1995;38(9)(Suppl):S240.

652. Kalunian KC, Moreland LW, Klashman DJ, Brion PH, Concoff AL, Myers S, Singh R, Ike RW, Seeger LL, Rich E and Skovron ML. Visually-guided irrigation in patients with early knee osteoarthritis: a multicenter randomized, controlled trial. Osteoarthritis and Cartilage. 2000;8(6):412-418.

653. Kaper BP, Smith PN, Bourne RB, Rorabeck CH and Robertson D. Medium term results of a mobile bearing total knee replacement. Clinical Orthopaedics and Related Research. 1999;No.367:201-209.

654. Karabatsos B, Mahomed NN and Maistrelli GL. Functional outcome of total knee arthroplasty after high tibial osteotomy. Canadian Journal of Surgery. 2002;45(2):116-119.

655. Karlsson J, Selin-Sjögren L and Wiklund I. A comparison of one generic and two disease-specific measurements of pain and physical function in patients with osteoarthrosis of the knee. Acta Orthop Scand. 1999;70(Suppl 287):56-57.

656. Katz JN, Losina E, Phillips CB, Mahomed N, Fossel AH, Barrett J, Lew RA, Lingard EA, Poss R, Miner A and Baron JA. Association between hospital volume of total hip replacement and patient-centered outcomes. Arthritis and Rheumatism. 2000;43(9)(Suppl):S106.

657. Katz JN, Phillips CB, Baron JA, Fossel AH, Mahomed NN, Barrett J, Lingard EA, Harris WH, Poss R, Lew RA, Guadagnoli E, Wright EA and Losina E. Association of hospital and surgeon volume of total hip replacement with functional status and satisfaction three years following surgery. Arthritis and Rheumatism. 2003;48(2):560-568.

658. Kaul H, Phillips CB, Fossel AH, Mahomed NN, Guadagnoli E, Lingard EA and Katz JN. Outcomes of primary and revision total hip replacement (THR) in a population-based cohort. Arthritis and Rheumatism. 1999;42(9)(Suppl):S266.

659. Keating EM, Ranawat CS and Cats-Baril W. Assessment of postoperative vigor in patients undergoing elective total joint arthroplasty: a concise patient- and caregiver -based instrument. Orthopedics. 1999;22(1)(Suppl.):s119-s128.

660. Kelley SS, Haden L and Wood ML. Cementation for femoral head osteonecrosis after collapse. American Academy of Orthopaedic Surgeons 71st Annual Meeting, March 10-14, 2004, San Francisco, Poster No. P085.

661. Kelly KD, Voaklander D, Kramer G, Johnston DWC, Redfern L, and Suarez-Almazor E. The impact of health status on waiting time for major joint arthroplasty. The Journal of Arthroplasty. 2000;15(7):877-883.

662. Kelly KD, Voaklander DC, Johnston DWC and Suarez-Almazor ME. Burden of illness while waiting for major joint arthroplasty. Arthritis and Rheumatism. 1999;42(9)(Suppl):S73.

663. Kelly KD, Voaklander DC, Johnston DWC and Suarez-Almazor ME. Equity in waiting times for major joint arthroplasty. Can J Surg. 2002;45(4):269-276 (editorial by M. Gross on p.247).

664. Kelly KD, Voaklander DC, Johnston DWC and Suarez-Almazor ME. Equity in waiting lists for major joint arthroplasty in a universal, publicly funded, health system. Arthritis and Rheumatism. 1999;42(9)(Suppl):S218.

665. Kelly KD, Voaklander DC, Johnston DWC, Newman SC and Suarez-Almazor ME. Change in pain and function while waiting for major joint arthroplasty. The Journal of Arthroplasty. 2001;61(3):351-359.

666. Kirkley A, Rampersaud R, Griffin S, Amendola A, Litchfield R and Fowler P. Tourniquet versus no tourniquet use in routine knee arthroscopy: A prospective, double-blind, randomized clinical trial. Arthroscopy. 2000;16(2):121-126.

667. Kocher MS, Steadman JR, Briggs KK, Sterett WI and Hawkins RJ. Reliability, validity, and responsiveness of the Lysholm knee scale for various chondral disorders of the knee. The Journal of Bone and Joint Surgery. 2004;86-A(6):1139-1145.

668. Kreibich DN,Vaz M, Bourne RB, Rorabeck CH, Kim P, Hardie R, Kramer J and Kirkley A. What is the best way of assessing outcome after total knee replacement? Clinical Orthopaedics and Related Research. 1996;331:221-225.

669. Lane JV and Howie CR. Outcome predictors in total knee replacement. Annals of the Rheumatic Diseases Abstracts 1999. XIV EULAR Congress, Glasgow, June 6-11, 1999, p.283.

670. Lane JV, Lingard EA and Howie CR. Integrated care pathways and outcome in total knee arthroplasty. The Journal of Bone and Joint Surgery [Br]. 2003;85-B(Suppl II):118.

671. Lattanzio P-J, Chess DG and MacDermid JC. Effect of the posterior cruciate ligament in knee-joint proprioception in total knee arthroplasty. The Journal of Arthroplasty. 1998;13(5):580-585.

672. Lau JTC, Koo JM,Davey JR, Schemitsch EH, Waddell JP and Mahomed NN. Determining the commonly used hip measures for total hip arthroplasty. The Canadian Orthopaedic Association 55th Annual Meeting, June 3-6, 2000, Edmonton, Alberta. Programme 2000, p.130.

673. Laupacis A, Bourne R, Rorabeck C, et al. The effect of elective total hip replacement on health-related quality of life. Journal of Bone and Joint Surgery [Am]. 1993;11:1619-1626.

674. Laupacis A, Bourne R, Rorabeck C, Feeny D, Tugwell P, Wong C. Comparison of total hip arthroplasty performed with and without cement. The Journal of Bone and Joint Surgery 2002;84-A(10):1823-8.

675. Lavernia CJ, Hernandez VH and Sierra R. Mental health and outcome in primary joint replacement. The Journal of Arthroplasty. 2002;17(2):257. American Association of Hip and Knee Surgeons, 11th Annual Meeting, Nov.9-11, 2001, Dallas, Texas.

676. Leitch KK, Weller P and Maistrelli G. Outcomes and economic costs of bilateral staged versus simultaneous total knee arthroplasty. The Canadian Orthopaedic Association 55th Annual Meeting, June 3-6, 2000, Edmonton, Alberta. Programme 2000, p.100.

677. Lieberman JR, Hawker G and Wright JG. Function in patients >55 years old. The Journal of Arthroplasty. 2001;16(7):901-904.

678. Lingard EA, Katz JN, Wright EA and Sledge CB. Predicting the outcome of total knee arthroplasty from preoperative status. The Journal of Bone and Joint Surgery [Br]. 2003;85-B(Suppl II):166.

679. Lingard EA, Katz JN, Wright EA, Sledge CB and Kinemax Outcomes Group. Predicting the outcome of total knee arthroplasty from preoperative status. Arthritis and Rheumatism. 2000;43(9)(Suppl):S216.

680. Lingard EA, Katz JN, Wright EA, Sledge CB and Kinemax Outcomes Group. Evaluating the impact of the process of care on the outcome of total knee replacement surgery: a multinational comparison. Arthritis and Rheumatism. 1999;42(9)(Suppl):S218.

681. Lingard EA, Katz JN, Wright EA, Sledge CB and Kinemax Outcomes Group. The relationship of patient expectations to outcome following total knee replacement surgery. Arthritis and Rheumatism. 2000;43(9)(Suppl):S217.

682. Lingard EA, Katz JN, Wright EA, Wright RJ, Sledge CB and Kinemax Outcomes Group. Validity and responsiveness of the Knee Society clinical rating system. Arthritis and Rheumatism. 2000;43(9)(Suppl):S217.

683. Lingard EA, Katz JN, Wright EA, Wright RJ, Sledge CB and Kinemax Outcomes Group. Does post discharge management of total knee replacement surgery effect outcome? Arthritis and Rheumatism. 2000;43(9)(Suppl):S215.

684. Lingard EA, Katz JN, Wright RJ, Wright EA, Sledge CB and the Kinemax Outcomes Group. Validity and responsiveness of the Knee Society clinical rating system in comparison with the SF-36 and WOMAC. The Journal of Bone and Joint Surgery. 2001;83-A(12):1856-1864.

685. Lingard EA, Miner AL, Wright EA, Sledge CB, Katz JN and Kinemax Outcomes Group. Knee range of motion after total knee arthroplasty: is this an important outcome measure. American Academy of Orthopaedic Surgeons 68th Annual Meeting, San Francisco, U.S.A. (Poster PE147).

686. Lingard EA, Wright EA, Sledge CB and the Kinemax Outcomes Group. Patient recall of preoperative pain and function 3 months following total knee arthroplasty. Arthritis and Rheumatism. 1999;42(9)(Suppl):S333.

687. Lingard EA, Wright EA, Sledge CB and the Kinemax Outcomes Group. Pitfalls of using patient recall to derive preoperative status in outcome studies of total knee arthroplasty. The Journal of Bone and Joint Surgery. 2001;83-A(8):1149-1156.

688. Lingard EA, Wright EA, Sledge CB and the Kinemax Outcomes Group. Pitfalls of using patient recall to derive preoperative status in outcome studies of total knee arthroplasty. American Academy of Orthopaedic Surgeons 68th Annual Meeting, Feb. 28- Mar. 4, 2001, San Francisco, U.S.A. (Paper 069).

689. Lingard EA, Wright RJ, Katz JN, Wright EA, Sledge CB and Kinemax Outcomes Group. Validity and responsiveness of the Knee Society clinical rating system. American Academy of Orthopaedic Surgeons 68th Annual Meeting, Feb. 28 - Mar. 4, 2001, San Francisco, U.S.A. (Poster PE148).

690. Lohmander S. Long-term outcome of total joint replacement for osteoarthritis. EULAR, Lisbon, Portugal, June 18-21, 2003.

691. MacDonald SJ, McCalden RW, Chess DG, Bourne RB, Rorabeck CH and Cleland D. Metal on metal versus metal on polyethylene in total hip arthroplasty – a prospective randomized clinical trial. American Academy of Orthopaedic Surgeons 70th Annual Meeting, Feb. 5-9, 2003, New Orleans.

692. MacDonald SJM, Rorabeck CH, MacDonald D and Clarke C. Outcomes following total arthroplasty: a randomized trial of intact versus sacrificed PCL. The Canadian Orthpaedic Association 55th Annual Meeting, June 3-6, 2000, Edmonton, Alberta. Programme 2000, p.181.

693. Machner A, Pap G and Awiszus F. Evaluation of quadriceps strength and voluntary activation after unicompartmental arthroplasty for medial osteoarthritis of the knee. Journal of Orthopaedic Research. 2002;20:108-111.

694. Machner A, Pap G, Krohn A, Rohkohl K and Awiszus F. Quadriceps muscle function after high tibial osteotomy for osteoarthritis of the knee. Clinical Orthopaedics and Related Research. 2002;No.399:177-183.

695. Mahomed N, Koo JM, Davey JR, Myerthall S, Gross AE, Davis A, Schemitsch EH and Waddell JP. Determinants of outcome following revision total knee arthroplasty. The Canadian Orthopaedic Association 56th Annual Meeting, London, Canada, June 1-4, 2001. Programme Book, p.108.

696. Mahomed N, Lau J, Davey JR, Davis A, Koo JM, Lin S, Schemitsch EH and Waddell JP. Determining the commonly used knee scores for total knee arthroplasty. The Canadian Orthopaedic Association 56th Annual Meeting, London, Canada, June 1-4, 2001. Programme Book, p.79.

697. Mahomed N, Wright JG, Hawker G, Davis AM and Badley E. RCT of home based vs inpatient rehab following TJA: functional outcomes and patient satisfaction. American Academy of Orthopaedic Surgeons 71st Annual Meeting, March 10-14, 2004, San Francisco. Paper presentation, Paper No. 154.

698. Mahomed NN, Arndt DC, McGrory BJ and Harris WH. The Harris Hip score. Comparison of patient self-report with surgeon assessment. The Journal of Arthroplasty. 2001;16(5):575-580.

699. Mahomed NN, Harris WH, Katz JN, Liang MH and Fossel AH. Imputation of WOMAC and SF-36 scores from the Harris Hip score. The Journal of Bone and Joint Surgery. 1999;81-B(Suppl.I):p.101.

700. Mahomed NN, Harris WH, Liang MH, Fossel AH and Katz JN. Imputation of WOMAC and SF-36 scores from the Harris Hip score. 66th Annual Meeting Proceedings of the American Academy of Orthopaedic Surgeons. February 4-8, 1999, Anaheim, CA, p.65.

701. Mahomed NN, Harris WH, Liang MH, Fossel AH and Katz JN. Imputation of WOMAC and SF-36 scores from the Harris Hip score. 66th Annual Meeting Proceedings of the American Academy of Orthopaedic Surgeons. February 4-8, 1999, Anaheim, CA, p.159.

702. Mahomed NN, Katz JN, Harris WJ, Philips C and Baron J. Mortality and complication rates following primary and revision total hip replacement (THR) are inversely related to hospital and surgeon procedure volume. Journal of Bone and Joint Surgery. 2001;83-B, Suppl. I:21.

703. Mahomed NN, Katz JN, Liang MH and Fossel AH. Role of patient expectations in predicting functional outcomes and satisfaction with total hip & knee arthroplasty. The Journal of Arthroplasty. 2000;15(2):p.257.

704. Mahomed NN, Katz JN, Liang MH and Fossel AH. The role of patient expectations in predicting functional outcomes and satisfaction with total hip and knee arthroplasty. The Journal of Bone and Joint Surgery. 1999;81-B(Suppl.I):p.106.

705. Mahomed NN, Koo JM, Levesque J, Lan S and Bogoch ER. Short-term outcomes of home based versus inpatient rehabilitation following total hip and knee arthroplasty. The Journal of Bone and Joint Surgery [Br]. 2000;82-B(Suppl.I):p.4.

706. Mahomed NN, Koo Soon Lin MJ, Levesque J, Lan S and Bogoch ER. Determinants and outcomes of inpatient versus home based rehabilitation following elective hip and knee replacement. The Journal of Rheumatology. 2000;27(7):1753-1758.

707. Mahomed NN, Liang MH, Cook EF, Daltroy LH, Fortin PR, Fossel AH and Katz JN. The importance of patient expectations in predicting functional outcomes after total joint arthroplasty. The Journal of Rheumatology. 2002;29:1273-1279.

708. Mahomed NN, Liang MH, Daltroy LH, Fortin PR and Katz JN. Patient expectations and total joint arthroplasty. Reply to above Letter to the Editor by Woolhead, Donovan and Dieppe. The Journal of Rheumatoloy. 2003;30(7):1657.

709. Mahon JL, Bourne RB, Rorabeck CH, Feeny DH, Stitt L and Webster-Bogaert S. Health-related quality of life and mobility of patients awaiting elective total hip arthroplasty: a prospective study. CMAJ. 2002;167(10):1115-1121.

710. March L, Cross M, Lapsley H, Tribe K, Courtenay B, Cross MJ, Pinczewski L, Ruff S, Coolican M, Neale M, Quain S, Robertson F, Walter W, Zicat B and Brooks P. 3 years down the track: patient expenditure and outcomes following primary and revision total hip & knee replacement. Osteoarthritis and Cartilage. 2002;10(Suppl A):S72-S73.

711. March L, Cross M, Tribe K, Lapsley H, Courtenay B and Brooks P. Cost of joint replacement surgery for osteoarthritis: the patient's perspectives. The Journal of Rheumatology. 2002;29(5):1006-1014.

712. March L, Cross M, Tribe K, Lapsley H, Coutenay B, Cross MJ, Pinczewski L and Brooks P. Patient expenditure and health outcomes following revision hip and knee joint replacement surgery for osteoarthritis. Internal Medicine Journal. 2003;33(Suppl):A3.

713. March LM, Cross M, Tribe KL, Lapsley HM, Courtenay BG, Cross MJ, Brooks PM For the Arthritis C.O.S.T. study project group: C. Cass, M. Coolican, M. Neil, L. Pinczewski, S. Quain, F. Robertson, S. Ruff, W. Walter and B. Zical. Two knees or not two knees? Patient costs and outcomes following bilateral and unilateral total knee joint replacement surgery for OA? Osteoarthritis and Cartilage. 2004;12(5):400-408.

714. Masri BA, Haddad F, Marko A, McEwen J, Garbuz DS and Duncan CP. Variations between estimated and actual pain-activity relationships in lower limb arthroplasty candidates: implications for improved assessment of patient outcomes. The Canadian Orthopaedic Association 55th Annual Meeting, June 3-6, 2000, Edmonton, Alberta. Programme 2000, p.172.

715. May CB and Gottschalk FAB. Varus supracondylar femoral osteotomy using a lateral approach. American Academy of Orthopaedic Surgeons 71st Annual Meeting, March 10-14, 2004, San Francisco, Poster P132.

716. McGrory BJ and Harris WH. Can the Western Ontario and McMaster Universities (WOMAC) Osteoarthritis index be used to evaluate different hip joints in the same patient? The Journal of Arthroplasty. 1996;11:841-844.

717. McGrory BJ, Freiberg AA, Shinar AA and Harris WH. Correlation of measured range of hip motion following total hip arthroplasty and responses to a questionnaire. The Journal of Arthroplasty.1996;11:565-571.

718. McGrory BJ, Shinar AA, Freiberg AA and Harris WH. Enhancement of the value of hip questionnaires by telephone follow-up evaluation. The Journal of Arthroplasty. 1997;12:340-343.

719. McKee MD. The effect of human recombinant bone morphogenic protein (rhBMP-7) on the healing of open tibial shaft fractures: results of a multi-centre, prospective, randomized clinical trial. American Academy of Orthopaedic Surgeons 70th Annual Meeting, Feb.5-9, 2003, New Orleans.

720. McNicholas MJ, Rowley DI, McGurty D, Adalberth T, Abdon P, Lindstrand A and Lohmander LS. Total meniscectomy in adolescence. The Journal of Bone and Joint Surgery [Br]. 2000;82-B(2):217-221.

721. Meek RMD, Dunlop DJ, Garbuz DS, Greidanus N, Masri BA, McGraw R and Duncan CP. Patient satisfaction and functional status after aseptic versus septic revision total knee arthroplasty using the PROSTALAC articulating spacer. American Academy of Orthopaedic Surgeons 70th Annual Meeting, Feb. 5-9, 2003, New Orleans.

722. Meek RMD, Greidanus N, Garbuz DS, Masri BA and Duncan CP. Effect of retaining a patellar prothesis on pain functional and satisfaction outcomes after revision total knee arthroplasty. American Academy of Orthopaedics Surgeons 70th Annual Meeting, Feb.5-9, 2003, New Orleans.

723. Mèzières M, Dougados M and Ayral X. Sensitivity to change of symptomatic outcome variables in knee osteoarthritis (OA). Annals of the Rheumatic Diseases Abstracts 1999. XIV EULAR Congress, Glasgow, June 6-11, 1999, p.156.

724. Minas T. Autologous chondrocyte implantation for focal chondral defects of the knee. Clinical Orthopaedics and Related Research. 2001;No.391S:S349-S361.

725. Minas T. Chondrocyte implantation in the repair of chondral lesions of the knee: economics and quality of life. The American Journal of Orthopedics. 1998; (11):739-744.

726. Miner AL, Lingard EA, Wright EA, Sledge CB and Katz JN. Knee range of motion after total knee arthroplasty: How important is this as an outcome measure? The Journal of Arthroplasty. 2003;18(3):286-294.

727. Miner AL, Lingard EA, Wright EA, Sledge CB, Katz JN and Kinemax Outcomes Group. Knee range of motion after total knee replacement is a poor determinant of outcome. Arthritis and Rheumatism. 2000;43(9)(Suppl):S219.

728. Mohler CG, Lantz BA and Velez-Duran M. Preoperative factors affecting quality of life after total knee arthroplasty. 66th Annual Meeting Proceedings of the American Academy of Orthopaedic Surgeons. February 4-8, 1999, Anaheim, CA, p.171.

729. Naylor CD, Williams JI and the Ontario Panel on Hip and Knee Arthroplasty. Primary hip and knee replacement surgery: Ontario criteria for case selection and surgical priority. Quality in Health Care. 1996;5:20-30.

730. Nilsdotter A, Aurell Y, Siösteen A-K, Lohmander S and Roos H. Radiographs and gender as predictors for outcome after total hip arthroplasty. Osteoarthritis and Cartilage. 2000;8(Suppl B):S81.

731. Nilsdotter A, Roos E, Westerlund J, Roos H and Lohmander S. Responsiveness of measures of pain and function after total hip replacement. Osteoarthritis and Cartilage. 2000;8(Suppl B):S81.

732. Nilsdotter A-K and Lohmander LS. Age and waiting time as predictors of outcome after total hip replacement for osteoarthritis. Rheumatology. 2002;41:1261-1267.

733. Nilsdotter A-K, Aurell Y, Siösteen A-K, Lohmander LS and Roos H. Radiographic stage of osteoarthritis or sex of the patient does not predict one year outcome after total hip arthroplasty. Annals of the Rheumatic Diseases. 2001;60:228-232.

734. Nilsdotter AK, Klassbo M, Lohmander SL and Roos EM. Hip Osteoarthritis Outcome Score in comparison to WOMAC – validity and responsiveness in total hip replacement. Arthritis and Rheumatism. 2002;46(9)(Suppl):S115.

735. Nilsdotter AK, Petersson IF, Roos EM and Lohmander LS. Predictors of unsatisfactory patient-relevant outcome after total hip replacement for osteoarthritis. Annals of the Rheumatic Diseases. 2002; 61 (Supplement). EULAR, June 12-15, Stockholm.

736. Nilsdotter A-K, Petersson IF, Roos EM and Lohmander LS. Similar pain but worse function compared with matched controls 3-5 years after total hip replacement for OA. Osteoarthritis and Cartilage. 2001;9(Suppl B):S48.

737. Nilsdotter A-K, Roos EM, Westerlund JP, Roos HP and Lohmander LS. Comparative responsiveness of measures of pain and function after total hip replacement. Arthritis Care and Research. 2001;45:258-262.

738. Nunez M, Nunez E, Quinto L, Rodriguez J, Macule F, Ma Segur J and Vilalta C. Health-related quality of life in patients with osteoarthritis on waiting list for total knee replacement: an interventional randomized study. Arthritis and Rheumatism. 2002; 46(9)(Suppl):S467.

739. Nunez M, Nunez E, Quinto LI, Macule F, Segur JM and Vilalta C. An approach to health-related quality of life and costs in patients on a knee replacement waiting list. Arthritis and Rheumatism. 2001;44(9)(Suppl):S184.

740. Parent E and Moffet H. Comparative responsiveness of locomotor tests and questionnaires used to follow early recovery after total knee arthroplasty. Archives of Physical Medicine and Rehabilitation. 2002;83:70-80.

741. Peloso PM, Cassidy JD and Carroll L. Who's on the queue? A study of patients waiting for hip and knee joint replacement surgery. Arthritis and Rheumatism. 1997;40(9)(Suppl):S311.

742. Peloso PM, Cassidy JD and Carroll L. What about the patient? Determinants of waiting times for hip and knee joint replacement surgery. Arthritis and Rheumatism. 1998;41(9)(Suppl):S88.

743. Peloso PM, Cassidy JD and Carroll L. Who's on the queue? A study of patients waiting for hip and knee joint replacement surgery. The Journal of Rheumatology. 1998;25(Suppl 52):4.

744. Pennington J, Hill R, Bayan A, Doyle T and Theis J-C. Clinical outcome of Duracon/PCA total condylar knee replacement: a 5 to 9 year follow-up. The Journal of Bone and Joint Surgery [Br]. 2003;85-B(Suppl III):203.

745. Poss R, Clark CR and Heckman JD. A concise format for reporting the longer-term follow-up status of patients managed with total hip arthroplasty. The Journal of Bone & Joint Surgery. 2001;83-A(12):1779-1781.

746. Poss R, Clark CR and Heckman JD. A concise format for reporting the longer-term follow-up status of patients managed with total knee arthroplasty. The Journal of Bone & Joint Surgery. 2002;84-A(7):1105-1106.

747. Price JS, Till SH, Bickerstaff DR, Bayliss MT and Hollander AP. Degradation of cartilage type II collagen precedes the onset of osteoarthritis following anterior cruciate ligament rupture. Arthritis and Rheumatism. 1999;42(11):2390-2398.

748. Psychoyios VN, Villanueva-Lopez F, Borven S, Crawford R, Hayes J and Murray D. Preoperative hip and knee arthritis severity under different healthcare delivery systems. The Journal of Bone and Joint Surgery [Br]. 2003;85-B(Suppl III):229.

749. Quintana JM, Arostegui I, Azkarate J, Ignacio Goenaga J, Elexpe X, Lotona J, Arcelay A. Evaluation of explicit criteria for total hip joint replacement. Journal of Clinical Epidemiology 2000;53:1200-8.

750. Radke S, Kirschner S, Seipel V, Rader C and Eulert J. Treatment of transient bone marrow oedema of the hip -- a comparative study. International Orthopaedics. 2003;27(3):149-152.

751. Ravaud P, Gerentes I, Taccoen A, Provvedini D. The rationale and design of PHARE, a new study on osteoarthritis. Osteoarthritis and Cartilage. 2003;11(Suppl A):S116.

752. Robertsson O and Dunbar MJ. Patient satisfaction compared with general health and disease-specific questionnaires in knee arthroplasty patients. The Journal of Arthroplasty. 2001;16(4):476-482.

753. Roos E. Knee injury and knee osteoarthritis - development, evaluation and clinical application of patient-relevant questionnaires. Acta Orthop Scand. 1999;Suppl.284:70\78.

754. Roos EM, Klässbo M and Lohmander LS. WOMAC Osteoarthritis Index. Reliability, validity and responsiveness in patients with arthroscopically assessed osteoarthritis. Paper III of Ewa Roos PhD Thesis.

755. Roos EM, Klässbo M and Lohmander LS. WOMAC Osteoarthritis Index. Reliability, validity, and responsiveness in patients with arthroscopically assessed osteoarthritis. Scandinavian Journal of Rheumatology. 1999;28:210-215.

756. Roos EM, Östenberg A, Roos H, Ekdahl C and Lohmander S. Long-term results of meniscectomy. Symptoms, function, and performance tests in patients with or without radiographic osteoarthritis compared to matched controls. Paper I of Ewa Roos PhD Thesis: Knee injury and knee osteoarthritis. Development, evaluation and clinical application of patient-relevant questionnaires. Dept. of Physical Therapy, Lund University, Sweden, 1999.

757. Roos EM, Roos HP and Lohmander S. WOMAC Osteoarthritis Index - additional dimensions for use in subjects with post-traumatic osteoarthritis of the knee. Osteoarthritis and Cartilage. 1999;7:216-221.

758. Roos EM, Roos HP, Ekdahl and Lohmander LS. Knee injury and Osteoarthritis Outcome Score (KOOS) - validation of a Swedish version. Scandinavian Journal of Medicine & Science in Sports 1998;8:439-448.

759. Roos EM, Roos HP, Johnson RJ, Ekdahl C and Lohmander LS. Knee injury and osteoarthritis outcome scores (KOOS): a reliable and valid outcome measure for meniscus injury, ACL injury, and osteoarthritis. Annals of the Rheumatic Diseases. 1996;55:693.

760. Roos EM, Roos HP, Lohmander LS, Ekdahl and Beynnon BD. Knee injury and Osteoarthritis Outcome Score (KOOS) - Development of a self-administered outcome measure. Journal of Orthopaedic and Sports Physical Therapy. 1998;28(2):88-96.

761. Roos EM, Roos HP, Ryd L and Lohmander LS. Substantial disability three months after arthroscopic partial meniscectomy. A prospective study of patient-relevant outcomes. Paper VI of Ewa Roos PhD Thesis.

762. Roos EW, Nilsdotter AK and Toksvig-Larsen S. Patients expectations suggest additional outcomes in total knee replacement. Arthritis and Rheumatism. 2002;46(9)(Suppl):S199.

763. Saleh KJ, Jaroszynski G, Woodgate E, Saleh L and Gross AE. Revision total hip arthroplasty with the use of structural acetabular allograft and reconstruction ring. A case series with a 10-year average follow-up. The Journal of Arthroplasty. 2000;15(9):951-958.

764. Saleh KJ, Radosevich DM, Kassim RA, Moussa M, Dykes D, Boltolfson H, Gloe TJ and Robinson H. Comparison of commonly used orthopaedic outcome measures using palm-top computers and paper surveys. Journal of Orthopaedic Research. 2002;20(6):1146-1151.

765. Salmon P and Hall GM. Postoperative fatigue is a component of the emotional response to surgery. Results of multivariate analysis. Journal of Psychosomatic Research. 2001;50:325-335.

766. Salmon P, Hall GM and Peerbhoy D. Influence of the emotional response to surgery on functional recovery during 6 months after hip arthroplasty. Journal of Behavioural Medicine. 2001;24(5):489-502.

767. Salmon P, Hall GM, Peerbhoy D, Shenkin A and Parker C. Recovery from hip and knee arthroplasty: patients' perspective on pain, function, quality of life, and well-being up to 6 months postoperatively. Archives of Physical Medicine and Rehabilitation. 2001;82:360-366.

768. Schemitsch EH, Bederman S and Waddell JP. How do traditional hip scoring systems correlate with validated disease-specific and general health outcome measures? 67th Annual Meeting Proceedings, American Academy of Orthopaedic Surgeons, March 15-19, 2000. Orlando, Florida, p.583.

769. Schemitsch EH, Bhandari M, Chan R, Kreder HJ, McKee MD, Stephen DJ and Whelan D. Validation of a scale to assess radiographic union of tibial shaft fractures treated with intramedullary nailing. The Canadian Orthopaedic Association 56th Annual Meeting, June 1-4, 2001, London, Canada, Programme Book, p.75.

770. Schick M, Stucki G, Rodriguez M, et al. Haemophilic arthroplasty: assessment of quality of life after total knee arthroplasty. Clin Rheumatol. 1999;18(6):468-472.

771. Singh GK, Jain R, Ahmad P, Ashish K and Singh K. Add on arthroscopic lavage (AL) in osteoarthritis (OA) knee: an open trial. Osteoarthritis and Cartilage. 2002;10(Suppl A):S76.

772. Söderman P and Malchau H. Is the Harris Hip score system useful to study the outcome of total hip replacement? Clinical Orthopaedics and Related Research. 2001;No.384:189-197.

773. Söderman P, Malchau H and Herberts P. Outcome measurements in total hip replacement surgery - a validation study on the Swedish National Hip Arthroplasty Register. Acta Orthop Scand. 1999;(Suppl.284)70:41.

774. Söderman P, Malchau H and Herberts P. Outcome of total hip replacement. A comparison of different measurement methods. Clinical Orthopaedics and Related Research. 2001;No.390:163-172.

775. Söderman P, Malchau H and Herberts P. The validity and reliability of Harris Hip score. Acta Orthop Scand. 1999;(Suppl.284)70:42.

776. Söderman P, Malchau H, Herberts P and Johnell O. Are the findings in the Swedish national total hip arthroplasty register valid? A comparison between the Swedish National Total Hip Arthroplasty Register, the National Discharge Register, and the National Death Register. The Journal of Arthroplasty. 2000;15(7):884-889.

777. Söderman P, Malchau H, Herberts P, Zügner R, Regnér H and Garellick G. Outcome after total hip arthroplasty. Part II. Disease-specific follow-up and the Swedish National Total Hip Arthroplasty Register. Acta Orthop Scand. 2001;72(2):113-119.

778. Söderman P. On the validity of the results from the Swedish National Total Hip Arthroplasty register. Acta Orthopaedica Scandinavica. 2000, Supplementum no. 296, vol. 71:1-33 [Thesis].

779. Soever LJ and Bajaj S. A community approach to total knee arthroplasty. Physiotherapy Canada. Winter 2001;53(Suppl):S22.

780. St. Pierre Y, Fortin PR, Ferland D, Katz JN, Tanzer M, Liang MH, Mahomed NN and Clarke AE. Total hip and knee replacement (THR/TKR): preoperative function and postoperative costs. Arthritis and Rheumatism. 1999;42(9)(Suppl):S328.

781. St. Pierre Y, Fortin PR, Katz JN, et al. Total hip and knee replacement: preoperative function and postoperative costs. Osteoarthritis and Cartilage. 2000;8(2):p.152.

782. *Stephens MAP, Druley JA and Zautra AJ. Older adults' recovery from surgery for osteoarthritis of the knee: psychosocial resources and constraints as predictors of outcomes. Health Psychology. 2002;21(4):377-383.*

783. *Stickles BJ, Phillips L, Owens BD, Anderson FA and Lieberman JR. Patient satisfaction after total hip arthroplasty and delta scores: a new relationship. American Academy of Orthopaedic Surgeons 68th Annual Meeting, Feb. 28 - Mar. 4, 2001, San Francisco, U.S.A. (Poster PE357).*

784. *Stickles BJ, Phillips L, Owens BD, Brox WT and Lanzer WL. Defining the relationship between obesity and total joint arthroplasty outcomes. American Academy of Orthopaedic Surgeons 68th Annual Meeting, Feb. 28 - Mar. 4, 2001, San Francisco, U.S.A. (Poster PE345).*

785. *Stulberg SD, Brander VA and Adams A. Predicting those at risk for greater total knee replacement pain: a prospective, observational study. American Academy of Orthopaedic Surgeons 70th Annual Meeting, Feb.5-9, 2003, New Orleans.*

786. *Swiontkowski MF. The outcomes movement in orthopaedic surgery: where we are and where we should go. The Journal of Bone and Joint Surgery. 1999;81-A:732-733.*

787. *Tanzer M, Fortin P, Clarke A, Joseph L, Liang MH, Ferland D, Phillips CB, Sledge C and Katz JN. Timing for total joint replacement affects patient outcomes. American Academy of Orthopaedic Surgeons 70th Annual Meeting, Feb. 5-9, 2003, New Orleans.*

788. *Tanzer M, Gollish JD, Leighton RK, Orrell KG, Giachino A, Wells GA and Shea B. Adjuvant calcium-phosphate coatings on proximally porous-coated total hip arthroplasties, an under utilized ally. American Academy of Orthopaedic Surgeons Annual Meeting, Feb. 13-17, 2002, Dallas.*

789. *Templeton JE, Goetz DD, Sullivan PM, Pederson DR, Johnston RC and Callaghan JJ. A benchmark for future total hip arthroplasty: results of Charnley THR at minimum 30-year follow-up. American Academy of Orthopaedic Surgeons 71st Annual Meeting, March 10-14, 2004, San Francisco, Poster P090.*

790. *Terek M. The cartilage thickness and functional status in patients with knee osteoarthritis. EULAR, Lisbon, Portugal, June 18-21, 2003.*

791. *Tile M, Borkhoff CM, Zaveri GR, Kreder HJ, Jiminez ML, Waddell JP and Stephen DJ. Outcome of total hip replacement following previous acetabular fracture versus total hip replacement for arthritis – a matched cohort study. American Academy of Orthopaedic Surgeons Annual Meeting, Feb. 13-17, 2002, Dallas.*

792. *Toye FM, Lamb SE, Barlow JH and Wright CC. The discriminatory validity of the New Zealand score for hip and knee surgery in a UK population. Annals of the Rheumatic Diseases. 2001;60(Suppl I):284.*

793. *Tribe K. Costs and outcomes of rehabilitation care for osteoarthritis hip and knee replacement patients. Internal Medicine Journal. 2003;33(Suppl):A8.*

794. *Tribe KL, Lapsley HM, March LM, Cross MJ, Courtenay B and Brooks PM. Impact of "locus of control" on outcomes of hip & knee replacement (THR & TKR) for osteoarthritis. Osteoarthritis and Cartilage. 2002;10(Suppl A):S73.*

795. *Tubach F, Baron G, Dougados M, Logeart I and Ravaud P. Predictive factors of total knee replacement due to primary osteoarthritis. Arthritis and Rheumatism. 2003;48(9)(Suppl):S77.*

796. *Tubach F, Baron G, Dougados M, Logeart I and Ravaud P. Predictive factors of total knee replacement due to primary osteoarthritis. EULAR, Lisbon, Portugal, June 18-21, 2003.*

797. *Tubach F, Baron G, Ravaud P, Logeart and Dougados M. Predictive factors of total hip replacement due to primary osteoarthritis. Arthritis and Rheumatism. 2003;48(9)(Suppl):S77.*

798. *Tubach F, Baron G, Ravaud P, Logeart I and Dougados M. Predictive factors of total hip replacement due to primary osteoarthritis. Osteoarthritis and Cartilage. 2003;11(Suppl A):S114-5.*

799. *Tubach F, Baron G, Ravaud P, Logeart I and Dougados M. Predictive factors of total hip replacement due to primary osteoarthritis. EULAR, Lisbon, Portugal, June 18-21, 2003.*

800. *Van Bergayk AB and Garbuz DS. Quality of life and sports-specific outcomes after Bernese periacetabular osteotomy. The Journal of Bone and Joint Surgery [Br]. 2002;84-B:339-343.*

801. *Van Bergeyk A and Garbuz D. Quality-of-life and sports-specific outcomes after periacetabular osteotomy. The Canadian Orthopaedic Association 56th Annual Meeting, June 1-4, 2001, London, Canada, Programme Book, p.105.*

802. *Wang A, Gilbey H, Ackland T and Morton A. Perioperative exercise programs improve functional recovery following total hip arthroplasty: a randomised prospective trial. American Academy of Orthopaedic Surgeons 68th Annual Meeting, Feb. 28 - Mar. 4, 2001, San Francisco, U.S.A. (Poster PE389).*

803. *Whelan D, Chan R, Bhandari M, Kreder HJ, Stephen DJ, McKee MD and Schemitsch EH. Validation of a scale to assess radiographic union of tibial shaft fractures treated with intramedullary nailing. American Academy of Orthopaedic Surgeons Annual Meeting, Feb. 13-17, 2002, Dallas.*

804. *Williams JI, Llewellyn-Thomas H, Arshinoff R, Young N and Naylor CD. The burden of waiting for hip and knee replacements in Ontario. Ontario Hip and Knee Replacement Project Team. Journal of Evaluation in Clinical Practice. 1997;3:59-68.*

805. *Williams JI, Young N and Arshinoff R. Comparing a disease-specific measure and a functional status measurement in assessing the waiting times and outcomes of hip and knee replacements. Quality of Life Research. 1994;3:99.*

806. *Witvrouw E, Victor J, Bellemans J, Rock B, Van Lummel R, van der Slikke R and Verdonk R. A correlation study of objective functionality and WOMAC in total knee arthroplasty. Knee Surgery, Sports Traumatology, Arthroscopy. 2002;10(6):347-351.*

807. *Wixted JJ, Grover NK, Anderson Jr. FA and Fitzgerald Jr RH. Clinical outcomes in orthopaedic surgery. The collaborative efforts of orthopaedic surgeons with industry. Bulletin Hospital for Joint Diseases. 1999;53(3):161-165.*

808. *Wolfe F, Nietfeld S, Hedrick R, McElrath T and Ross K. Length of hospital stay and complications of total joint replacement are unrelated to functional status, pain, or psychological variables. Arthritis and Rheumatism. 1993;36(9)(Suppl):S59.*

809. *Woodgate IG, Saleh KJ, Jaroszynski G, Agnidis Z, Woodgate MM and Gross AE. Minor column structural acetabular allografts in revision hip arthroplasty. Clinical Orthopaedics and Related Research. 2000;No.371:75-85.*

810. *Woolhead G, Donovan J and Dieppe P. Patient expectations and total joint arthroplasty. The Journal of Rheumatology. 2003;30(7):1656-1657. [Letter to the editor]*

811. *Wright JG, Coyte PC, Hawker GA, Croxford R and Bombardier C. What determines geographic area variation in knee replacement utilization? The Journal of Bone and Joint Surgery. 1999;81-B(Suppl.I):p.103.*

812. *Young N, Williams JI, Wright JG and Piexto G. Rating difficulty of functioning in patients undergoing hip or knee replacements: comparing the WOMAC and physical functioning (SF-36) scales. Quality of Life Research. 1994;3:100.*

Miscellaneous Studies (WOMAC).

813. *Arden NK, Nevitt MC, Williams E, Gore LR, Lane NE for the Study of Osteoporotic Fractures. The incidence of radiographic hip osteoarthritis in elderly white women. Annals of the Rheumatic Diseases Abstracts 1999. XIV EULAR Congress, Glasgow, June 6-11, 1999, p.258.*

814. *Axford JM, Ross F, Triggs E and Victor C. Osteoarthritis of the knee causes appreciable disability in females associated with depression and inability to cope. Arthritis and Rheumatism. 1999;42(9)(Suppl):S142.*

815. *Axford JS, Ross F, Triggs E and Victor C. Osteoarthritis of the knee causes appreciable disability in females associated with depression and inability to cope. Arthritis and Rheumatism. 2000;43(9)(Suppl):S221.*

816. *Axford JS, Ross F, Triggs E and Victor C. Osteoarthritis of the knee causes appreciable disability in females associated with depression and inability to cope. Rheumatology. 2000;39(Abstracts Suppl.1):p.126.*

817. *Axford S, Triggs E, Ross F, Cadbury H and Victor C. Lack of efficacy of a structured care package for osteoarthritis of the knee. Arthritis and Rheumatism. 1998;41(9)(Suppl):S199.*

818. *Ayral X, Pickering EH, Woodworth TG, Loose LD, MacKillip N and Dougados M. Synovitis is not correlated with the level of symptomatic severity in painful knee osteoarthritis patients. Annals of the Rheumatic Diseases. 2002; 61 (Supplement). EULAR, June 12-15, Stockholm.*

819. *Bae S-C, Lee H-S, Seo J-C, Kim T-H, Jun J-B, Jung S-S, Lee I-H, Yoo D-H and Kim S-Y. Psychometric evaluation of Korean WOMAC and Lequesne index for clinical research. Arthritis and Rheumatism. 2000;43(9)(Suppl):S110.*

820. *Bae S-C, Lee H-S, Yun HR, Kim T-H, Yoo D-H and Kim SY. Cross-cultural adaptation and validation of Korean Western Ontario and McMaster Universities (WOMAC) and Lequesne Osteoarthritis Indices for clinical research. Osteoarthritis and Cartilage. 2001;9:746-750.*

821. *Bain O, Glusser G, Graif M and Halperin N. Assessment of osteoarthritis of the knee: comparison of different imaging techniques. EULAR, Lisbon, Portugal, June 18-21, 2003.*

822. *Baker K, McLaughlin S, Li W, LaValley M, Goggins J and Felson D. Quadriceps weakness and its association with change in pain and disability in knee osteoarthritis: a 30-month longitudinal, observational study. Arthritis and Rheumatism. 2002;46(9)(Suppl):S414.*

823. *Battle-Gualda E, Esteve-Vives J, Piera Riera MC, Hargreaves R and Cutts J. Translation and adaptation to Spanish of the WOMAC questionnaire specific for knee and hip osteoarthritis. Revista Espanola de Reumatologia 1999;26(2):*

824. *Bellamy N, Goldstein LD and Tekanoff RA. Continuing medical education-driven skills acquisition and impact on improved patient outcomes in family practice setting. The Journal of Continuing Education in the Health Professions. 2000;20:52-61.*

825. *Bellamy N, Roszko P and Hall D. How important is flare in the design of NSAID trials in the treatment of osteoarthritis (OA)? Arthritis and Rheumatism. 2000;43(9)(Suppl):S221.*

826. *Birrell F, Lingard E, Khan A, Francis R, Peaston R and McCaskie A. The prevalence and distribution of osteoporosis in patients with end-stage hip osteoarthritis (OA): evidence for OA mitigating local bone loss. EULAR 2004, June 9-12, Berlin, Germany.*

827. *Bolognese JA, Ehrich EW and Schnitzer TJ. Precision of composite measures of osteoarthritis efficacy in comparison to that of individual endpoints. The Journal of Rheumatology. 2001;28(12):2700-2704.*

828. Bombardier C, Maguire L, Li L, Maetzel A, Jansz G and The CHAP Team. Characteristics of patients with osteoarthritis seen by family physicians: results of a telephone survey in Ontario. The Journal of Rheumatology. 2001;28(6):1416.

829. Botha-Scheepers SA, Riyazi N, Kroon HM, Slagboom E, Meulebelt I, Roendaal FR, Breedveld FC and Kloppenburg M. Pain and disability in the lower extremities in patients with osteoarthritis (OA) at multiple sites are associated with lumbar spine OA, the number of OA joints, range of motion and psychological factors. EULAR 2004, June 9-12, Berlin, Germany.

830. Bradley JD, Heilman DK and G'Sell P. Do psychological factors "predict" response to tidal lavage (TL) and sham lavage (SL) in knee osteoarthritis (KOA)? Arthritis and Rheumatism. 2000;43(9)(Suppl):S337.

831. Bradley JD, Heilman DK, Katz BP, G'Sell P, Wallick JE and Brandt KD. Tidal irrigation as treatment for knee osteoarthritis. A sham-controlled, randomized, double-blinded evaluation. Arthritis and Rheumatism. 2002;46(1):100-108.

832. Braun J, Alten R, Burmester G, Ebner W, Gomor B, Grasedyck K, Hammer M, HantzscheiH, Hermann J, Kallwasser J, Kellner H, Krause A, Mielke F, Schneider M, Tony H, Zeidler H, Beck E, Baumgaertner M and Sieper J. Efficacy of sulfasalazine in undifferentiated spondyloarthritis and early ankylosing spondylitis – a multicenter randomized controlled trial. EULAR 2004, June 9-12, Berlin, Germany.

833. Brenner SS, Klotz U, Alscher DM, Mais A, Lauer G, Schweer H, Seyberth HW, Fritz P and Bierbach U. Osteoarthritis of the knee – clinical assessments and inflammatory markers. Osteoarthritis and Cartilage. 2004;12(6):469-475.

834. Briganti MP, Di Bari M, Mannoni A, Ferrucci L, Serni U, Masotti G and Marchionni N. Determinants of disability in community-dwelling older persons with knee osteoarthritis: the ICARe Dicomano study. Annals of the Rheumatic Diseases. 2000;59(Suppl).

835. Brooks LO, Rolfe M, Cheras PA and Myers SP. Reliability, validity and responsiveness of a simple index for measurement of treatment effects in longitudinal clinical trials of osteoarthritis treatments. Osteoarthritis and Cartilage. 2002;10(Suppl A):S63-S64.

836. Brooks LO, Rolfe MI, Cheras PA and Myers SP. The Comprehensive Osteoarthritis Test: a simple index for measurement of treatment effects in clinical trials. The Journal of Rheumatology. 2004;31(6):1180-1186.

837. Bruce B and Fries JF. Longitudinal comparison of the Health Assessment Questionnaire (HAQ) and the Western Ontario McMasters Universities Osteoarthritis Index (WOMAC). Arthritis and Rheumatism. 2003;48(9)(Suppl):S640.

838. Bruyere O, Ethgen O, Lejeune E, Giacovelli G, Richy F, Seidel L, Henrotin Y, Rovati L, and Reginster J-Y. Clinical severity of knee osteoarthritis poorly predicts long-term radiographic outcomes. Arthritis and Rheumatism. 2002;46(9)(Suppl):S153.

839. Bruyere O, Henrotin Y, Honore A, Rovati L, Giacovelli G, Seidel L and Reginster J-Y. Tibial subchondral sclerosis and femoral osteophytes are linked to symptomatic and structural severity of knee osteoarthritis. Osteoarthritis and Cartilage. 2001;9(Suppl B):S22.

840. Bruyere O, Honore A, Rovati LC, Giacovelli G, Henrotin YE, Seidel L and Reginster J-YL. Radiologic features poorly predict clinical outcomes in knee osteoarthritis. Scandinavian Journal of Rheumatology. 2002;31:13-16.

841. Byers Kraus V, Gell N and Blumenthal JA. The effect of chronic exercise on circulating biomarkers in individuals with musculoskeletal disease. Clinical Exercise Physiology 1999;1(1):

842. Cerejo R, Dunlop DD, Cahue S, Channin D, Song J, Sharma L. The influence of alignment on risk of knee osteoarthritis progression according to baseline stage of disease. Arthritis and Rheumatism 2002;46(10):2632-2636.

843. Chalsson CE, Gale DR, Gale ME, Totterman S and Felson DT. Prevalence of bone marrow lesions and their association with knee OA pain and severity. Arthritis and Rheumatism. 1999;42(9)(Suppl):S293.

844. Chassany O, Boureau F, Liard F, Bertin P, Serrie A, Marchand S, Keddad K and Jolivet-Landreau I. Clinical study on the impact of training in pain management in general medical practice on the evolution of pain in patients with osteoarthritis of the lower limbs. Arthritis and Rheumatism. 2003;48(9)(Suppl):S632-S633.

845. Chassany O, Boureau F, Liard F, Bertin P, Serrie A, Marchand S, Keddad K and Jolivet-Landreau I. Clinical study on the impact of training in pain management in general medical practice on the evolution of pain in patients with osteoarthritis of the lower limbs. EULAR, Lisbon, Portugal, June 18-21, 2003.

846. Choquette D, Bellamy N and Raynauld JP. A French-Canadian version of the WOMAC Osteoarthritis Index. Arthritis and Rheumatism. 1994;37(9)(Suppl):S226.

847. Christensen R, Astrup A and Bliddal H. Weight loss: the treatment of choice for knee osteoarthritis? A randomised trial. EULAR 2004, June 9-12, Berlin, Germany.

848. Christgau S, Henrotin Y, Henriksen D, Rovati L, Collette J, Deroisy R, Christiansen C and Reginster J-Y. Cartilage degradation in knee osteoarthritis patients with elevated levels of urinary collagen type II C-telopeptide fragments. Arthritis and Rheumatism. 2000;43(9)(Suppl):S274.

849. Cibere J, Bellamy N, Thorne A, Esdaile JM, McGorm KJ, Chalmers A, Huang S, Peloso P, Shojania K, Singer J, Wong H and Kopec J. Reliability of the knee examination in osteoarthritis. Effect of standardization. Arthritis and Rheumatism. 2004;50(2):458-468.

850. Cibere J, Bellamy N, Thorne A, Esdaile SM, McGorm K, Chalmers A, Huang S, Peloso P, Singer J, Shojania K and Kopec J. Examination in osteoarthritis reliability and standardization of the knee. Osteoarthritis and Cartilage. 2002;10(Suppl A):S62-S63.

851. Cicuttini F, Wluka A, Wang Y and Stuckey S. The determinants of change in patella cartilage volume in osteoarthritic knees. The Journal of Rheumatology. 2002;29(12): 2615-2619.

852. Clarke S, Andrews S, Glew D, Dunmall P, Locke V and Hollingworth XP. Radiological hip and knee osteoarthritis (OA) is more frequent in experienced unilateral amputees. Arthritis and Rheumatism. 2000;43(9)(Suppl):S398.

853. Clarke S, Duddy J, Wakeley C, Nickols G, Sharif M, Watt I and Kirwan JR. Bony swelling predicts positive bone scintigraphy in mild to moderate knee osteoarthritis (OA). Arthritis and Rheumatism. 2000;43(9)(Suppl):S337.

854. Coleman S, Conroy H, Rowley E, Fabling M, Briffa K, Carroll G, Prince R and McQuade J. Long term effects of education in the primary health care setting using behaviour modification, and specific exercises in patients with osteoarthritis of the knee: reduced pain, improved mobility and improved quality of life. Arthritis and Rheumatism. 2002;46(9)(Suppl):S198-S199.

855. Conaghan PG, Grainger A, Hensor E, Keenan A, Thomas C, Rhodes L and Emery P. The relationship between pain and MRI synovitis in osteoarthritis of the knee. EULAR 2004, June 9-12, Berlin, Germany.

856. Concoff A, Singh R, Klashman D, et al. A clinical algorithm for identifying occult crystalline disease in patients with knee osteoarthritis. Arthritis and Rheumatism. 1997;40(9)(Suppl):S239.

857. Creamer P, Lethbridge-Cejku M and Hochberg M. Factors associated with pain severity reporting in knee osteoarthritis (OA). Arthritis and Rheumatism. 1997;40(9)(Suppl):S237.

858. Creamer P, Lethbridge-Cejku M and Hochberg M. Factors associated with functional impairment in knee osteoarthritis (OA). Arthritis and Rheumatism. 1997;40(9)(Suppl):S237.

859. Creamer P, Lethbridge-Cejku M and Hochberg M. Where does it hurt? Pain localization in osteoarthritis of the knee. Arthritis and Rheumatism 1997;40(9)(Suppl):S237.

860. Creamer P, Lethbridge-Cejku M and Hochberg MC. Determinants of pain severity in knee osteoarthritis: effect of demographic and psychosocial variables using 3 pain measures. The Journal of Rheumatology. 1999;26(8):1785-1792.

861. Creamer P, Lethbridge-Cejku M and Hochberg MC. Factors associated with functional impairment in symptomatic knee osteoarthritis. Rheumatology. 2000;39:490-496.

862. Creamer P, Lethbridge-Cejku M and Hochberg MC. Where does it hurt? Pain localization in osteoarthritis of the knee. Osteoarthritis and Cartilage. 1998;6:318-323.

863. Creamer P, Lethbridge-Cejku, Hochberg MC, Zeng O, Bolognese J and Ehrich E. Individual radiographic features (IRFs) are not associated with severity of pain and physical disability in patients with knee osteoarthritis (KOA). Arthritis and Rheumatism. 1998;41(9)(Suppl):S87.

864. Cunningham J, Sinkfield D, Groner K, Williams D, Gracely RR, Mohan N, Cupps TR, Chaplin W, and Clauw DJ. The effect of "central" factors on knee symptoms. Arthritis and Rheumatism. 2000;43(9)(Suppl):S336.

865. Do J-H, Kang H-J, Kim C-H, Ko H-J, Ok Jung Y, Min D-J, Kim W-U, Min J-K, Park S-H, Lee S-H, Cho C-S, and Kim H-Y. Correlation of knee joint pain with capsular distension and length of osteophytes on ultrasonography in patients with osteoarthritis. Arthritis and Rheumatism. 2002;46(9)(Suppl):S151.

866. Domsic RT and Saltzman CL. Ankle osteoarthritis scale. Foot & Ankle International. 1998;19(7):466-471.

867. Dougados M, Mézières M and Ayral X. Sensitivity to change of symptomatic outcome variables in knee osteoarthritis (OA). Arthritis and Rheumatism. 1998;41(9)(Suppl):S199.

868. Dougados M. Response criteria in OA trials. Annals of the Rheumatic Diseases. 2000;59(Suppl 1).

869. Edworthy SM, Devins GM, and the Patient Education Study Group. Improving medication adherence through patient education distinguishing between appropriate and inappropriate utilization. The Journal of Rheumatology. 1999;26(8):1793-1801.

870. Escobar A, Quintana JM, Bilbao A, Azkárate J and Güenaga JI. Validation of the Spanish version of the WOMAC questionnaire for patients with hip or knee osteoarthritis. Clinical Rheumatology. 2002;21:466-471.

871. Etchells E. A safe, effective adjunctive treatment for arthritis, and pretty much every other condition you'll ever treat. The Journal of Rheumatology. 1999;26(8):1647-1649.

872. Evcik D and Kuru Y. Correlation of mechanical and anatomic axis of knee joint with functional capacity of the patients with osteoarthritis. EULAR, Lisbon, Portugal, June 18-21, 2003.

873. Faucher M, Poiraudeau S, Lefevre-Colau MM, Rannou F, Fermanian J and Revel M. Assessment of the test-retest reliability and construct validity of a modified Lequesne index in knee osteoarthritis. Joint Bone Spine 2003;70:520-525. [Also published in French in Revue du Rhumatisme 2003;70:1105-111. Évaluation de la reproductibilité et de la validité de construit d'une forme modifiée de l'indice algofunctionnel de Lequesne dans l'arthrose du genou.

874. Faucher M, Poiraudeau S, Lefevre-Colau MM, Rannou F, Fermanian J and Revel M. Algo-functional assessment of knee osteoarthritis: comparison of the test-retest reliability and construct validity of the WOMAC and Lequesne indexes. Osteoarthritis and Cartilage 2002;10:602-610.

875. Fitzgerald GK, Piva SR, Irrgang JJ, Bouzubar F and Starz TW. Quadriceps activation failure as a moderator of the relationship between quadriceps strength and physical function in individuals with knee osteoarthritis. Arthritis Care and Research. 2004;51(1):40-48.

876. Fitzgerald K, Piva SR, Irrgang JJ, Bouzubar F and Starz TW. Quadriceps activation failure as a moderator of the relationship between quadriceps strength and physical function in individuals with knee osteoarthritis. Arthritis and Rheumatism. 2002;46(9)(Suppl):S198.

877. Fleischmann R, Moore A and Della Casa Alberighi O. Osteoarthritis Research Society International (OARSI) responder rates in patients with knee osteoarthritis receiving lumiracoxib, celecoxib or placebo. Arthritis and Rheumatism. 2003;48(9)(Suppl):S78-S79.

878. Flugsrud GB, Espehaug B, Havelin LI, Nordsletten L and Meyer HE. Body mass index and physical activity as risk factors for primary osteoarthritis. A cohort study in 50,034 persons. Annals of the Rheumatic Diseases. 2001;60(Suppl I):77-78.

879. Fortin PR, Abrahamowicz M, Badley E, King LE, du Berger R, Chang E and Poole R. Discriminatory ability and reliability of serological biomarker for osteoarthritis. Arthritis and Rheumatism. 2003;48(9)(Suppl):S684.

880. Gagnier P and Hwang E. The use of the OARSI responder index as a tool to assess dose response relationships in OA efficacy trials. Annals of the Rheumatic Diseases. 2002;61 (Supplement). EULAR, June 12-15, Stockholm.

881. Garnero P, Piperno M, Christgau S, Delmas PD and Vignon E. Biochemical markers of bone, cartilage and synovium tissue turnover in patients with knee osteoarthritis: correlation with disease activity and joint destruction. Annals of the Rheumatic Diseases. 2000;59(Suppl 1).

882. Garnero P, Piperno M, Ginoyts E, Christgau S, Delmas PD and Vignon E. Cross sectional evaluation of biochemical markers of bone, cartilage, and synovial tissue metabolism in patients with knee osteoarthritis: relations with disease activity and joint damage. Annals of the Rheumatic Diseases. 2001;60:619-626.

883. Garnero P, Piperno M, Ginoyts E, Christgau S, Delmas PD and Vignon E. Biochemical markers of bone, cartilage and synovium tissue turnover in patients with knee osteoarthritis: correlation with disease activity and joint destruction. Arthritis and Rheumatism. 2000;43(9)(Suppl):S337.

884. Gentelle-Bonnassies S, Le Claire P, Mezieres M, Ayral X and Dougados M. Comparison of the responsiveness of symptomatic outcome measures in knee osteoarthritis. Arthritis Care and Research. 2000;13(5):280-285.

885. Glazier RH, Hawker GA, Wright JG, Coyte PC, Williams JI, Harvey B, Wilkins AL and Badley EM. Patient and provider factors related to arthritis care in a community setting. Arthritis and Rheumatism. 2002;46(9)(Suppl):S410-S411.

886. Goggins J, Niu J, McLaughlin S, LaValley M, Li W, Gale D, Gale M and Felson D. Body Mass Index (BMI) and its longitudinal effect on disability and pain in persons with knee osteoarthritis (KOA). Arthritis and Rheumatism. 2002;46(9)(Suppl):S198.

887. Gokoglu F, Yorgancioglu Z, Ustun N and Findikoglu G. Evaluation of hand strength and functional ability in patients with generalized osteoarthritis. EULAR, Lisbon, Portugal, June 18-21, 2003.

888. Guermazi M, Poiraudeau S, Yahia M, Mezganni M, Fermanian J, Elleuch H and Revel M. Translation, adaptation and validation of the Western Ontario and McMaster Universities osteoarthritis index (WOMAC) for an Arab population: the Sfax modified WOMAC. Osteoarthritis and Cartilage. 2004;12(6):459-468.

889. Günther KP, Stürmer T, Sauerland S, Zeissig I, Sun Y, Kessler S, Scharf HP, Brenner H and Puhl W. Prevalence of generalised osteoarthritis in patients with advanced hip and knee osteoarthritis: The Ulm Osteoarthritis Study. Annals of the Rheumatic Diseases. 1998;57:717-723.

890. Guymer EK, Williams DA, Cunningham J, Sinkfield D, Gracely RH and Clauw DJ. Factors related to physical function outcome measures in non-inflammatory rheumatic disease. Arthritis and Rheumatism. 2001;44(9)(Suppl):S356.

891. Guymer EK, Williams DA, Cunningham J, Sinkfield D, Gracely RH and Clauw DJ. Predictors of 50 foot walk test time in individuals with non-inflammatory rheumatic disease. Arthritis and Rheumatism. 2001;44(9)(Suppl):S140.

892. Halbert JA, Silagy CA, Finucane P Withers RT and Hamdorf PA. Recruitment of older adults for a randomized, controlled trial of exercise advice in a general practice setting. Journal of the American Geriatrics Society. 1999;47:477-481.

893. Hashimoto H, Hanyu T, Sledge CB and Lingard EA. Validation of a Japanese patient-derived outcome scale for assessing total knee arthroplasty: comparison with Western Ontario and McMaster Universities osteoarthritis index (WOMAC). Journal of Orthopaedic Science. 2003;8(3):288-293.

894. Henderson CJ, Young DM, Conn DI. Characterization of food intake and physical activity patterns in urban African American women with knee osteoarthritis (OA). Arthritis and Rheumatism. 2001;44(9)(Suppl):S156.

895. Hill J, Thorpe RF and Bird HA. Comparison of outcome from a nurse-led clinic compared to standard outpatient care. Arthritis and Rheumatism. 2002;46(9)(Suppl):S555.

896. Hill J. The outcome of OA patients from a nurse-led clinic. EULAR 2004, June 9-12, Berlin, Germany.

897. Hinman RS, Bennell KL, Metcalf BR and Crossley KM. Balance impairments in individuals with symptomatic knee osteoarthritis: a comparison with matched controls using clinical tests. Rheumatology. 2002;41(12):1388-1394.

898. Hinman RS, Bennell KL, Metcalf BR and Crossley KM. Delayed onset of quadriceps activity and altered knee joint kinematics during stair stopping in individuals with knee osteoarthritis. Archives Physical Medicine and Rehabilitation. 2002;83:1080-1086.

899. Hinman RS, Bennell KL, Metcalf BR, Crossley KM. Temporal activity of vastus medialis obliqius and vastus lateralis in symptomatic knee osteoarthritis. American Journal of Physical Medicine and Rehabilitation 2002;81:684-690.

900. Hochberg MC, Bausell B, Frick K, Lee WL and Berman B. Validation of the revised OMERACT-OARSI responder index: responders have better overall health status than nonresponders. Osteoarthritis and Cartilage. 2003;11(Suppl A):S6-7.

901. Hochberg MC, Bausell B, Frick K, Steinwachs D and Berman B. Validation of the OARSI responder criteria: responders have better overall health status than nonresponders. Osteoarthritis and Cartilage. 2002;10(Suppl A):S1.

902. Hochberg MC, Bujak D, Frick K, Steinwachs D, Bausell B and Berman B. Validation of the OARSI responder criteria for knee osteoarthritis: responders have better overall health status than nonresponders. Arthritis and Rheumatism. 2002;46(9)(Suppl):S157.

903. Hunter DJ, March LM and Sambrook PN. The association of altered subchondral architecture with pain in knee osteoarthritis. Arthritis and Rheumatism. 2001;44(9)(Suppl):S140.

904. Hunter DJ, March LM and Sambrook PN. The association of cartilage volume with knee pain. Osteoarthritis and Cartilage. 2003;11:725-729.

905. Hurwitz DE, Ryals AB, Case JP, Block JA and Andriacchi TP. The knee adduction moment during gait in subjects with knee osteoarthritis is more closely related with static alignment than radiographic disease severity, toe out angle and pain. Journal of Orthopaedic Research. 2002;20:101-107.

906. Jacobs M, Kudaimi T, Byrd D, Katz B and Brandt K. Arthrogenous muscle inhibition (AMI) in subjects with knee osteoarthritis (KOA). Arthritis and Rheumatism. 1998;41(9)(Suppl):S86.

907. Jinks C, Jordan K and Croft P. Measuring the population impact of knee pain and disability with the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC). Pain. 2002;100, Issues 1-2, November:55-64.

908. Jinks C, Jordan K, Nio Ong B and Croft PR. The influences on health care use in older adults with knee pain. Arthritis and Rheumatism. 2002;46(9)(Suppl):S468.

909. Jinks C, Lewis M and Croft P. Validity and reliability of a brief screening tool for knee pain in primary care. Rheumatology. 2000;39 (Abstracts Suppl. 1):p.127.

910. Jinks C, Lewis M, Ong BN and Croft P. A brief screening tool for knee pain in primary care. Validity and reliability. Rheumatology. 2001;40:528-536.

911. Kaçar C, Urhan S, Arikan V and Bütün B. C-reactive protein levels in osteoarthritis: association with clinical and radiological findings. Annals of the Rheumatic Diseases. 2002; 61 (Supplement). EULAR, June 12-15, Stockholm.

912. Kahan A and Lieu PL. Relationship between OARSI response criteria and patients global assessment of treatment efficacy in knee osteoarthritis (OA). Osteoarthritis and Cartilage. 2001;9(Suppl B):S20.

913. Katz JN, Solomon DH, Schaffer JL, Horsky J, Burdick E and Bates DW. Outcomes of care and resource utilization among patients with knee or shoulder disorders treated by general internists, rheumatologists, or orthopedic surgeons. The American Journal of Medicine. 2000;109:28-35.

914. Katz P, Lee FH and Bahrt K. Depression and patients' ratings of the effectiveness of pain medications for knee osteoarthritis. Arthritis and Rheumatism. 2003;48(9)(Suppl):S608.

915. Khraishi MM, Curran V, Jong M, McComisky A, Rahman P and Ravalia M. Distance education strategies: can we improve outcomes of rural arthritis patients? Annals of the Rheumatic Diseases. 2002; 61 (Supplement). EULAR, June 12-15, Stockholm.

916. Kim Y-J, Burstein D, Gray ML, Jaramillo D and Millis MB. Assessment of dysplasic hip joint using delayed gadolinium enhanced MRI of cartilage (dGEMRIC). Osteoarthritis and Cartilage. 2001;9(Suppl B):S6.

917. Klässbo M, Larsson E and Mannevik E. Hip disability and osteoarthritis outcome score. An extension of the Western Ontario and McMaster Universities Osteoarthritis Index. Scandinavian Journal of Rheumatology. 2003;32:46-51.

918. Klässbo M, Larsson E, Mannevik E and Roos E. Hip dysfunction and osteoarthritis outcome score (HOOS) an extension of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC). Osteoarthritis and Cartilage. 2000;8(Suppl B):S76.

919. Klässbo M, Larsson G and Harm-Ringdahl K. Outcome of the 'hip school' - education program for patients with hip dysfunction. Osteoarthritis and Cartilage. 2000;8(Suppl B):S76.

920. Klässbo M, Larsson G and Harms-Ringdahl K. Promising outcome of a hip school for patients with hip dysfunction. Arthritis Care and Research. 2003;49(3):321-327.

921. Kosinski M, Bayliss MS, Bjorner JB, Ware JE, Laliberte K, Sullivan E, Vogel SM and Straus WI. Validation of the Osteoarthritis Impact Assessment. Arthritis and Rheumatism. 2001;44(9)(Suppl):S137.

922. Kristofferson H, Qvistgaard E, Terslev L, Torp-Pedersen S, Holm C, Danneskiold-Samsoe B and Bliddal H. Colour doppler activity as sign of synovitis in knee osteoarthritis. Annals of the Rheumatic Diseases. 2002; 61 (Supplement). EULAR, June 12-15, Stockholm. (AB0464 and SAT0257 listed twice in abstracts).

923. Krocson S and Theiler R. Is the WOMAC a useful tool as knee OA outcome measure in clinical practice: a survey. Osteoarthritis and Cartilage. 2000;8(Suppl B):S77.

924. Kuzmanova SI, Atanassov A, Andreev S, Klinkanov K and Solakov P. Arthroscopic synovectomy, debridement and tidal lavage of the knee joint in osteoarthritis – 3 years' follow-up study. EULAR, Lisbon, Portugal, June 18-21, 2003.

925. Lamb SE, Guralnik JM, Buchner DM, Ferrucci LM, Hochberg MC, Simonsick EM, Fried LP, for the Women's Health and Aging Study. Factors that modify the association between knee pain and mobility limitation in older women: the Women's Health and Aging Study. Annals of the Rheumatic Diseases. 2000;59:331-337.

926. Lamb SE, Toye F, Barker KL and Barrington S. Changes in measures of pain and function in intervention studies of severe knee osteoarthritis: comparison to patient perceptions. Annals of the Rheumatic Diseases. 2001;60(Suppl I):285.

927. Lane NE, Oehlert J, Bloch D and Fries JF. What factors predict which individuals with knee OA treated with NSAIDS will agree to randomization to acetaminophen. Arthritis and Rheumatism.1998;41(9)(Suppl):S181.

928. LaValley M, McAlindon TE, Evans S, Chaisson CE and Felson DT. Problems in the development and validation of questionnaire-based screening instruments for ascertaining cases with symptomatic knee osteoarthritis. The Framingham Study. Arthritis and Rheumatism. 2001;44(5):1105-1113.

929. Leardini G, Salaffi F, Montanelli R, Gerzeli S, Colangelo I and Canesi B. A multicentre study of annual costs of knee osteoarthritis in Italy. Arthritis and Rheumatism. 2001;44(9)(Suppl):S313.

930. Lee E, Shin K, Lee Y, Kang S, Baek H, Kim H and Song Y. Validity and reliability of Korean version of Lequesne Index. Annals of the Rheumatic Diseases. 2002; 61 (Supplement). EULAR, June 12-15, Stockholm.

931. Leffondre K, Abrahamowicz M, Hawker GA, Badley EM, Regeasse A, McCusker J and Belzile E. Longitudinal patterns of change in disability in osteoarthritis. Arthritis and Rheumatism. 2003;48(9)(Suppl):S664.

932. Leisen J, Duncan H and Kardasz S. Clinical similarities between black and white females with osteoarthritis (OA) of the knee. Arthritis and Rheumatism. 1993;36(5)(Suppl):R26.

933. Leisen J, Duncan H and Kardasz S. Clinical similarities between black and white females with osteoarthritis (OA) of the knee. Arthritis and Rheumatism. 1993;36(9)(Suppl):S167.

934. Leisen J, Duncan H and Kardasz S. The clinical significance of osteoarthritis (OA) of the hand in patients with OA knee. Arthritis and Rheumatism. 1993;36(9)(Suppl):S59.

935. Lolson J, Duncan H, Lubetsky M and Kardasz S. The utility of an outcome questionnaire in a short term clinical setting. Arthritis and Rheumatism. 1993;36(5)(Suppl):R12.

936. Lequesne MG, Méry C, Samson M and Marty M. Comparison between the WOMAC and the Lequesne indices in patients with knee and hip osteoarthritis. [Letter to the Editor]. Osteoarthritis and Cartilage. 1998;6:441-442.

937. Leveille SG, Ling S, Hochberg MC, Resnick HE, Bandeen-Roche KJ, Won A and Guralnik JM. Widespread musculoskeletal pain and the progression of disability in older disabled women. Annals of Internal Medicine. 2001;135:1038-1046.

938. Lindeman E, Leffers P, Spaans F, Drukker J, Reulen J, Kerckhoffs M and Köke A. Strength training in patients with myotonic dystrophy and hereditary motor and sensory neuropathy: a randomized clinical trial. Archives of Physical Medicine and Rehabilitation. 1995;76:612-620.

939. Ling SM, Fried LP, Garrett ES, Fan M-Y, Rantanen T and Bathon JM. Knee osteoarthritis compromises early mobility function: The Women's Health and Aging Study II. The Journal of Rheumatology. 2003;30(1):114-120.

940. Lingard E, Birrell F, Khan A, Francis R, Peaston R and McCaskie A. The prevalence and distribution of osteoporosis in patients with end-stage knee osteoarthritis (OA): more evidence for OA mitigating local bone loss. EULAR 2004, June 9-12, Berlin, Germany.

941. Lissin D, Kothari M, Sloffort M, Miaux Y, von Ingersleben G, Li J, Stevens RM and Peterfy C. Selecting x-ray inclusion/exclusion criteria for multicenter studies of osteoarthritis. Osteoarthritis and Cartilage. 2000;8(Suppl B):S76-79.

942. Loeser RF, Jr., Messier SP, Morgan TM, Chua S, Lenz ME and Thonar E. Effects of an exercise and dietary intervention on serum biomarkers in overweight and obese adults with osteoarthritis of the knee. Arthritis and Rheumatism. 2003;48(9)(Suppl):S698-S699.

943. Lord J, Victor C, Littlejohns P, Ross FM and Axford JS. Economic evaluation of a primary care-based education programme for patients with osteoarthritis of the knee. Health Techno Assess. 1999;3(23):1-55.

944. Maguire L, Li L, Bombardier C, Maetzel A, Jansz G and the CHAP Team. Characteristics of patients with osteoarthritis seen by family physicians: results of a telephone survey in Ontario. Arthritis Health Professions Association 18th Annual Scientific Professional Conference and Annual General Meeting, Toronto, Canada, May 4, 2001. [Poster Presentation]

945. Manek NJ, Wolfe F, Fries JF and Lane NE. Are osteoarthritic patients using alternative medications really different? Arthritis and Rheumatism. 1999;42(9)(Suppl):S406.

946. Mazzuca SA, Brandt KD and Surber J. Reduction of severe joint pain increases joint space width (JSW) in standing extended-view x-rays of patients with knee osteoarthritis (OA). Arthritis and Rheumatism. 2001;44(9)(Suppl):S155.

947. Mazzuca SA, Brandt KD, Katz BP, Caffrey HM, Ragozzino L and G'Sell P. Effects of self-care education for patients with knee osteoarthritis (OA) on prescribing practices of primary care physicians (PCPs). Arthritis and Rheumatism. 2001;44(9)(Suppl):S207.

948. Mazzuca SA, Brandt KD, Katz BP, et al. Differences in nonpharmacologic management of knee osteoarthritis (OA) by rheumatologists (RHs) and primary care physicians (PCPs). Arthritis and Rheumatism. 1994;37(9)(Suppl):S229.

949. Mazzuca SA, Brandt KD, Katz BP, et al. Rheumatologists (RHs) offer more comprehensive care for osteoarthritis (OA) of the knee than family physicians (FPs) or general internists (GIMs). But does it matter? Arthritis and Rheumatism. 1995;38(9)(Suppl):S239.

950. Mazzuca SA, Brandt KD, Katz BP, et al. Comparison of general internists, family physicians, and rheumatologists managing patients with symptoms of osteoarthritis of the knee. Arthritis Care and Research. 1997;10:289-299.

951. Mazzuca SA, Brandt KD, Katz BP, Olin M, Ragozzino LR and Caffrey HM. Controlled evaluation of a self-care intervention for patients with knee osteoarthritis (OA) in a health maintenance organization (HMO). Arthritis and Rheumatism. 2000;43(9)(Suppl):S331.

952. Mazzuca SA, Brandt KD, Lane KA and Katz BP. Knee pain reduces joint space width in conventional standing anteroposterior radiographs of osteoarthritic knees. Arthritis and Rheumatism. 2002;46(5):1223-1227.

953. Meddeb N, Hamza S, Zitouni M, Hamdi W, Mseddi N, Siala M, Chahed M, Dziri C, Chabene M, Laadhar L, Aouadni H, Makni S and Sellami S. Is MMP 13 and TIMP 1 intra articular fluid levels predictive of evolution of knee osteoarthritis? Osteoarthritis and Cartilage. 2003;11(Suppl A):S73.

954. Mikesky A, Perkins S, Caffroy H, Mazzuca S and Brandt K. Radiographic knee osteoarthritis (OA) with and without joint pain. Arthritis and Rheumatism. 2001;44(9)(Suppl):S138.

955. Mitchell H, Carr AJ, Virji A, Rogers A, Tovey D and Scott DL. Problems with the implementation of guidelines in primary care: absentee patients. Annals of the Rheumatic Diseases Abstracts 1999. XIV EULAR Congress, Glasgow, June 6-11, 1999, p.288.

956. Mohan N, Guymer EK, Williams DA, Cunningham J, Sinkfield D, Gracely RH and Clauw DJ. Varying personal beliefs in locus of pain control are associated with different outcomes in patients with chronic pain. Arthritis and Rheumatism. 2001;44(9)(Suppl):S66.

957. Nevitt M, Felson D, Potorfy C, Wildy K, Ling S, Lane N, Newman A, Carbone L and Harris T. Inflammation markers (CRP, TNF-a, IL-6) are not associated with radiographic or MRI findings of knee OA in the elderly: the Health ABC study. Arthritis and Rheumatism. 2002;46(9)(Suppl):S372-373.

958. Nevitt M, Goodpaster B, Felson D, Newman A, Wildy K, Harris F, Kritchevsky S, Simonsick E and Harris T. High prevalence of symptomatic knee OA (SXKOA) in elderly black women: role of obesity and quadriceps weakness. Arthritis and Rheumatism. 2001;44(9)(Suppl):S375.

959. Nevitt MC, Arden NK, Lane NE, Gore LR, Hochberg MC for the Study of Osteoporotic Fractures Research Group. Incidence of radiographic changes of hip OA in elderly white women. Arthritis and Rheumatism. 1998;41(9)(Suppl):S181.

960. O'Malley KJ, Souchek J, Byrne M and Suarez-Almazor M. The Joint-specific Multidimensional Assessment of Pain (J-MAP): comparison with the WOMAC. Arthritis and Rheumatism. 2003;48(9)(Suppl):S488.

961. O'Malley KJ, Suarez-Almazor M, Aniol J, Richardson P, Kuykendall DH, Moseley Jr. JB and Wray NP. Joint-Specific Multidimensional Assessment of Pain (J-MAP): Factor structure, reliability, validity, and responsiveness in patients with knee osteoarthritis. The Journal of Rheumatology. 2003;30(3):534-543.

962. O'Reilly SC, Jones A, Muir KR and Doherty M. Quadriceps strength and depression are important factors in knee pain and associated disability. Annals of the Rheumatic Diseases. 1996;55:693.

963. O'Reilly SC, Muir KR and Doherty M. Radiographic features of knee osteoarthritis in the community: the relation with pain, disability, and quadriceps strength. Annals of the Rheumatic Diseases.1996;55:693.

964. Oatis CA, Lockard MA, Michener LA and Goldovsky A. A comparison of WOMAC scores, gait variables and stiffness and damping characteristics in individuals with knee OA with and without crepitus. Arthritis and Rheumatism. 1997;40(9)(Suppl):S31.

965. Pal Y-C, Rymer WZ, Chang RW and Sharma L. Effect of age and osteoarthritis on knee proprioception. Arthritis and Rheumatism. 1997;40(12):2260-2265.

966. *Pavelka K, Bruyere O, Rovati LC, Olejarova M, Giacovelli G and Reginster J-Y. Relief in mild-to-moderate pain is not a confounder in joint space narrowing assessment of full extension knee radiographs in recent osteoarthritis structure-modifying drug trials. Osteoarthritis and Cartilage. 2003;11(10):730-737.*

967. *Pavelka K, Rovati LC, Gatterova J, Bruyere O, Deroisy R, Machacek S, Giacovelli G, Gonzalez C and Reginster JY. Pain relief is not a confounder in joint space narrowing assessment of full extension knee radiographs. Osteoarthritis and Cartilage. 2002;10(Suppl A):S16-S17.*

968. *Pavelka K, Rovati LC, Gatterova J, Bruyere O, Deroisy R, Machacek S, Giacovelli G, Gonzalez C and Reginster JY. Pain relief is not a confounder in joint space narrowing assessment of full extension knee radiographs. Annals of the Rheumatic Diseases. 2002; 61 (Supplement). EULAR, June 12-15, Stockholm.*

969. *Payne MWC. The effect of viscosupplementation on proprioception in the osteoarthritic knee. M.Sc. Thesis. The University of Western Ontario, London, Canada. 1998.*

970. *Perell K, Guerrero E, Fang M, Bragas M, Miyamoto E and Peralta M. Comparison of gait kinetics across disease severity in knee osteoarthritis subjects. Arthritis and Rheumatism. 1998;41(9)(Suppl):S86.*

971. *Persson L-O, Berglund K and Sahlberg D. Psychological factors in chronic rheumatic diseases - a review. Scandinavian Journal of Rheumatology. 1999;28:137-144.*

972. *Pessis E, Drapé J-L, Ravaud P, Chevrot A, Dougados M and Ayral X. Assessment of progression in knee osteoarthritis: results of a 1 year study comparing arthroscopy and MRI. Osteoarthritis and Cartilage. 2003;11:361-369.*

973. *Pham T, van der Heijde D, Altman RD, Anderson JJ, Bellamy N, Hochberg M, Simon L, Strand V, Woodworth T and Dougados M. OMERACT-OARSI initiative: Osteoarthritis Research Society International set of responder criteria for osteoarthritis clinical trials revisited. Osteoarthritis and Cartilage. 2004;12(5):389-399.*

974. *Piva SR, Fitzgerald GK, Irrgang J, Bouzubar F and Starz TW. Get up and go test in patients with knee osteoarthritis. Arch Phys Med Rehabil 2004;85:284-289.*

975. *Piva SR, Fitzgerald K, Irrgang JJ, Bouzabar F and Starz TW. Reliability and validity of the get up and go test in patients with knee osteoarthritis. Arthritis and Rheumatism. 2002;46(9)(Suppl):S228.*

976. *Pope J, Stevens A, Fenlon D and Mahon J. A randomized trial of conventional treatment versus N of 1 trials in osteoarthritis (OA). Cost effectiveness of treatment (Rx) strategies. Arthritis and Rheumatism. 1998;41(9)(Suppl):S79.*

977. *Ramey DR, Fries JF, Singh G and Lane NE. Outcome assessment in osteoarthritis (OA): WOMAC or HAQ? Arthritis and Rheumatism. 1998;41(9)(Suppl.):S223.*

978. *Raynauld J, Pelletier J, Beaudoin G, Berthiaume M-J, De Guise JA, Bloch DA, Altman RD, Hochberg M, Cline G, Buckland-Wright JC, Meyer J and Martel-Pelletier J. A two-year study in osteoarthritis patients following the progression of the disease by magnetic resonance imaging using a novel quantification imaging system. Annals of the Rheumatic Diseases. 2002; 61 (Supplement). EULAR, June 12-15, Stockholm.*

979. *Raynauld J-P, Kauffmann C, Godbout B, Beaudoin G, Berthiaume M-J, DeGuise J, Gagnon R, Bloch D, Altman R, Martel-Pelletier J, Choquette D, Cline G, Meyer J and Pelletier J-P. Knee osteoarthritis progression evaluated by magnetic resonance imaging and a novel quantification software tool. Arthritis and Rheumatism. 2000;43(9)(Suppl):S399.*

980. *Raynauld JP, Kauffmann C, Godbout B, Beaudoin G, Berthiaume MJ, DeGuise J, Gagnon R, Bloch D, Altman A, Martel-Pelletier J, Cline G, Meyer J and Pollotior JP. Knee osteoarthritis progression evaluated by magnetic resonance imaging and a novel quantification software tool. Osteoarthritis and Cartilage. 2000;8(Suppl B):S82-83.*

981. *Raynauld J-P, Pelletier J-P, Beaudoin G, Berthiaume M-J, de Guise JA, Bloch DA, Altman RD, Hochberg M, Cline G, Buckland-Wright C, Meyer J, and Martel-Pelletier J. A two-year study in osteoarthritis patients following the progression of the disease by magnetic resonance imaging using a novel quantification imaging system. Arthritis and Rheumatism. 2002;46(9)(Suppl):S150.*

982. *Reginster J-Y, Bruyere O, Rovati L, Honore A, Giacovelli G, Seidel L and Henrotin Y. Radiologic features poorly predict clinical outcomes in knee osteoarthritis. Osteoarthritis and Cartilage. 2001;9(Suppl B):S49.*

983. *Reginster J-Y, Collette J, Rovati L, Deroisy R, Honore A, Bruyere O, Seidel L and Henrotin Y. Biochemical markers of bone and cartilage remodeling are poorly correlated with clinical and radiological severity of knee osteoarthritis. Arthritis and Rheumatism. 2000;43(9)(Suppl):S339.*

984. *Reginster JY, Henrotin Y, Bruyere O, Gonzalez C, Giacovelli G and Rovati LC. Joint space narrowing is poorly correlated with symptoms worsening in knee osteoarthritis: results from the prospective follow-up of 106, placebo-treated patients for three years. Annals of the Rheumatic Diseases. 2002; 61 (Supplement). EULAR, June 12-15, Stockholm.*

985. *Robinson DB, Wood S, and Canvin JMG. Effect of a patient eduction program on WOMAC scores in lower limb osteoarthritis. The Journal of Rheumatology. 1999;26(7):1627.*

986. *Ross F, Victor C, Triggs E and Axford JS. Pain is an important predictor of outcome when a structured care package for osteoarthritis knee treatment is implemented. Arthritis and Rheumatism. 1999;42(9)(Suppl):S296.*

987. Salaffi F, Leardini G, Canesi B, Mannoni A, Fioravanti A, Caporali R, Lapadula G, et al. Reliability and validity of the Western Ontario and McMaster Universities (WOMAC) Osteoarthritis Index in Italian patients with osteoarthritis of the knee. Osteoarthritis and Cartilage. 2003;11(8):551-560.

988. Santos FP, Sampaio J, Cobrado N, Parente M, Paiva V, Serras M, Martins J, Mesquita F and Patto J. Body composition parameters influence in the knee osteoarthritis. EULAR 2004, June 9-12, Berlin, Germany.

989. Sautner J, Rintelen B, Wolf J and Leeb BF. SACRAH (Score for Assessment and quantitation of Chronic Rheumatic Affections of the Hands). Arthritis and Rheumatism. 2000;43(9)(Suppl):S223.

990. Scanzello CR, Pearle AD, Di Carlo EF, Ghelman B, Sculco TP and Crow MK. Elevated plasma CRP levels in osteoarthritis (OA) are associated with infiltrating synovial T-lymphocytes as well as increased pain and disability. Arthritis and Rheumatism. 2002;46(9)(Suppl):S495.

991. Schiavon F, Marotta H, Maffei P, Martini C and Menegazzo C. Acromegalic arthropathy: an Italian study in 20 patients. Annals of Rheumatic Diseases. 2001; 60(Suppl I):240.

992. Scott-Lennox JA, McLaughlin-Miley C, Lennox RD, Bohlig AM, Cutler BL, Yan C and Jaffe M. Stratification of flare intensity identifies placebo responders in a treatment efficacy trial of patients with osteoarthritis. Arthritis and Rheumatism. 2001. 44(7):1599-1607.

993. Sharma L, Cahue S, Song J, Hayes K and Dunlop D. The influence of joint laxity, other mechanical factors, and psychosocial factors on physical functioning over 3 years in knee osteoarthritis. Arthritis and Rheumatism. 2002;46(9)(Suppl):S577.

994. Sharma L, Cahue S, Song J, Hayes K, Pai Y-C and Dunlop D. Physical functioning over three years in knee osteoarthritis. Role of psychosocial, local mechanical, and neuromuscular factors. Arthritis and Rheumatism. 2003;48(12):3359-3370.

995. Sharma L, Cooper Z, Dupuis J and Chang RW. Gender predicts patient satisfaction with conservative therapy in knee osteoarthritis. Arthritis and Rheumatism. 1995;38(9)(Suppl):S382.

996. Sharma L, Dunlop DD, Cahue S, Song J and Hayes KW. Quadriceps strength and osteoarthritis progression in malaligned and lax knees. Annals of Internal Medicine. 2003;138:613-619.

997. Sharma L, Hayes KW, Felson DT, Buchanan TS, Kirwan-Mellis G, Lou C, Pai Y-C and Dunlop D. Does laxity alter the relationship between strength and function in knee osteoarthritis? Arthritis and Rheumatism. 1998;41(9)(Suppl):S352.

998. Sharma L, Hayes KW, Felson DT, Buchanan TS, Kirwan-Mellis G, Lou C, Pai Y-C and Dunlop DD. Does laxity alter the relationship between strength and physical function in knee osteoarthritis? Arthritis and Rheumatism. 1999;42(1):25-32.

999. Sharma L, Hayes KW, Pai Y-C, et al. Determinants of functional impairment in knee OA patients with patellofemoral involvement. Arthritis and Rheumatism. 1997;40(9)(Suppl):S238.

1000. Sharma L, Lou C, Felson DT, Dunlop DD, Kirwan-Mellis G, Hayes KW, Weinrach D and Buchanan TS. Laxity in healthy and osteoarthritic knees. Arthritis and Rheumatism. 1999;42(5):861-870.

1001. Sharma L, Pai Y-C, Holtkamp K and Rymer WZ. Is knee joint proprioception worse in the arthritic knee versus the unaffected knee in unilateral knee osteoarthritis? Arthritis and Rheumatism. 1997;40(9):1518-1525.

1002. Sharma L. The relationship between specific radiographic features and functional status in knee osteoarthritis. Arthritis and Rheumatism. 1996;39(9)(Suppl):S172.

1003. Shepstone L and Cline G. Statistical properties of OARSI type responder definitions. Osteoarthritis and Cartilage. 2003;11(Suppl A):S113. [Not a WOMAC reference but a measurement reference.]

1004. Shepstone L and Cline GA. Statistical properties of OARSI type responder definitions. Arthritis and Rheumatism. 2003;48(9)(Suppl):S399. [Not a WOMAC reference but a measurement reference.]

1005. Shepstone L, Koutantji M and Hand C. Patterns of osteoarthritic knee pain in primary care - a pilot study. Osteoarthritis and Cartilage. 2000;8(Suppl B):S85.

1006. Slatkowsky-Christensen B, Bellamy N and Kvien TK. Examinations of the Norwegian version of AUSCAN - a disease specific measure of hand osteoarthritis (OA).Annals of the Rheumatic Diseases. 2001;60(Suppl I):232.

1007. Slatkowsky-Christensen B, Kvien TK and Bellamy N. Performance of the Norwegian version of AUSCAN - disease specific measure of hand osteoarthritis (OA). Scandinavian Journal of Rheumatology. 2002;31:51 (Annual meeting of the Norwegian Society for Rheumatology, Oslo, 28 November 2001).

1008. Slemenda C, Brandt KD, Heilman DK, et al. Quadriceps weakness and osteoarthritis of the knee. Annals of Internal Medicine. 1997;127:97-104.

1009. Slemenda C, Heilman D, Brandt K, Braunstein E and Mazzuca S. A prospective study of knee pain and function in a random sample of elderly. Arthritis and Rheumatism. 1997;40(9)(Suppl):S216.

1010. Slemenda C, Heilman DK, Brandt KD, Katz BP, Mazzuca SA, Braunstein EM and Byrd D. Reduced quadriceps strength relative to body weight. A risk factor for knee osteoarthritis in women? Arthritis and Rheumatism. 1998;41(11):1951-1959.

1011. Slemenda C, Hellman D, Brandt K, Braunstein E and Mazzuca S. Weight, adiposity and quadriceps weakness as predictors of progression of knee osteoarthritis (OA). Arthritis and Rheumatism. 1997;40(9)(Suppl):S239.

1012. Söderman P and Malchau H. Validity and reliability of Swedish WOMAC osteoarthritis index. A self-administered disease-specific questionnaire (WOMAC) versus generic instruments (SF-36 and NHP). Acta Orthop Scand. 2000;71(1):39-46.

1013. Solomon DH, Katz JN, Schaffer JL, Burdick E, Horsky J and Bates DW. Rheumatologists provide different care for musculoskeletal conditions. Arthritis and Rheumatism. 1999;42(9)(Suppl):S219.

1014. Stucki G, Meier D, Stucki S, et al. Evaluation einer deutschen Version des WOMAC (Western Ontario and McMaster Universities) Arthroseindex. Zeitschrift für Rheumatologie. 1996;55:40-49.

1015. Stucki G, Sangha O, Stucki S, et al. Comparison of the WOMAC Osteoarthritis Index and the Lequesne Index in patients with knee and hip osteoarthritis. Arthritis and Rheumatism. 1996;39(9)(Suppl):S260.

1016. Stucki G, Sangha O, Stucki S, Michel BA, Tyndall A, Dick W and Theiler R. Comparison of the WOMAC (Western Ontario and McMaster Universities) osteoarthritis index and a self-report format of the self-administered Lequesne-Algofunctional index in patients with knee and hip osteoarthritis. Osteoarthritis and Cartilage. 1998;6:79-86.

1017. Stürmer T, Günther K-P and Brenner H. Obesity, overweight and patterns of osteoarthritis: the Ulm Osteoarthritis Study. Journal of Clinical Epidemiology. 2000; 53:307-313.

1018. Sun Y, Sturmer T, Gunther KP and Brenner H. Reliability and validity of clinical outcome measurements of osteoarthritis of the hip and knee - a review of the literature. Clinical Rheumatology. 1997;16(2):185-198.

1019. Tallon D, Chard J and Dieppe P. Exploring the priorities of patients with osteoarthritis of the knee. Arthritis Care and Research. 2000;13(5):312-319.

1020. Tamm AE, Hansen Ü and Tamm AO. Radiographic knee osteoarthritis in an Estonian symptomatic population sample in age 34-54 years. Osteoarthritis and Cartilage. 2003;11(Suppl A):S114.

1021. Theiler R, Sangha O, Schaeren S, Michel BA, Tyndall A, Dick W and Stucki G. Superior responsiveness of the pain and function sections of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) as compared to the Lequesne-algofunctional index in patients with osteoarthritis of the lower extremities. Osteoarthritis and Cartilage. 1999;7:515-519.

1022. Thorstensson CA, Roos EM, Jacobsson LTH, Boegard TL and Petersson IF. Worse physical function predicts appearance of radiographic knee osteoarthritis and more symptoms five years later in middle aged subjects with chronic knee pain. Arthritis and Rheumatism. 2001;44(9)(Suppl):S155.

1023. Thumboo J, Chew L-H and Soh C-H. Validation of the Western Ontario and McMaster University Osteoarthritis Index in Asians with osteoarthritis in Singapore. Osteoarthritis and Cartilage. 2001;9:440-446.

1024. Turchin DC, Schemitsch EH, McKee MD and Waddell JP. Do foot injuries significantly affect the functional outcome of multiply injured patients? Journal of Orthopaedic Trauma. 199;13(1):1-4.

1025. Victor C, Ross F, Triggs E and Axford JS. Lack of efficacy with a structured care package for osteoarthritis of the knee. Annals of the Rheumatic Diseases Abstracts 1999. XIV EULAR Congress, Glasgow, June 6-11, 1999, p.124.

1026. Villanueva I, del Mar Guzman M and Navarro F. Sensitivity and specificity of the Osteoarthritis Research Society International (OARSI) response criteria initiative (RCI): the effect of three different approaches for measuring pain. Arthritis and Rheumatism. 2001;44(9)(Suppl):S137.

1027. Villanueva I, del Mar Guzman M, Javier Toyos F, Ariza-Ariza R and Navarro F. Relative efficiency and validity properties of a visual analogue vs a categorical scaled version of the Western Ontario and McMaster Universities Osteoarthritis (WOMAC) Index: Spanish versions. Osteoarthritis and Cartilage. 2004;12:225-231.

1028. Villanueva I, Navarro F, Toyos FJ, Ariza R and Guzmán MM. Validation of a Spanish version of the WOMAC (Western Ontario and McMaster Universities Osteoarthritis Index). Arthritis and Rheumatism. 1997;40(9)Suppl:S110.

1029. Villanueva I, Toyos FJ, Ariza-Ariza R, Guzman MM and Navarro F. Validity of the Osteoarthritis Research Society International (OARSI) response criteria initiative (RCI): a comparison with patient's perceived level of improvement. Arthritis and Rheumatism. 2000;43(9)(Suppl):S224.

1030. Walker CRC, Myles C, Nutton R and Rowe P. Movement of the knee in osteoarthritis - the use of electrogoniometry to assess function. The Journal of Bone and Joint Surgery. 2001;83-B:195-198.

1031. Wigler I, Neumann L and Yaron M. Validation study of a Hebrew version of WOMAC in patients with osteoarthritis of the knee. Clinical Rheumatology. 1999;18(5):402-405.

1032. Wildy KS, Nevitt MC, , Kwoh CK, Lane N, Krishnaswami S, Peterfy CG, Kritchevsky S, Harris T, Zaim S, Guermazi , Felson D and Newman AB. MRI findings associated with knee pain: analysis of discordant knee pairs in health ABC. Arthritis and Rheumatism. 2002;46(9)(Suppl):S148.

1033. Wolfe F and Mosser J. Disease severity, treatment, demographics and the risk of hospitalization in RA. Arthritis and Rheumatism. 2001;44(9)(Suppl):S379.

1034. *Wolfe F and Pincus T. Listening to the patient. A practical guide to self-report questionnaires in clinical care. Arthritis and Rheumatism. 1999;42(9):1797-1808.*

1035. *Wolfe F, Flowers N and Anderson J. Comparative socio-economic, clinical and outcome status of minority and non-minority patients with rheumatoid arthritis. Arthritis and Rheumatism. 2000;43(9)(Suppl):S139.*

1036. *Wolfe F, Flowers N and Anderson J. Prednisone and functional disability, not DMARDS or biologics, explain increased infection rates in rheumatoid arthritis (RA). Arthritis and Rheumatism. 2000;43(9)(Suppl):S403.*

1037. *Wolfe F, Flowers N and Anderson J. The key importance of household income among demographic and socio-economic variables in RA. Arthritis and Rheumatism. 2000;43(9)(Suppl):S139.*

1038. *Wolfe F, Michaud K, Kahler K and Omar M. The Short Arthritis Assessment Scale (SAS): a brief assessment questionnaire for rapid evaluation of arthritis severity in clinical and research use. Arthritis and Rheumatism. 2003;48(9)(Suppl):S400.*

1039. *Wolfe F, Nlotfeld S, Hedrick R, McElrath T and Ross K. Gender difference and care for osteoarthritis: Differential severity for OA of the hip and knee as a function of gender. Arthritis and Rheumatism. 1993;36(9)(Suppl):S177.*

1040. *Woodworth TG, Mackillop N, Edwards M, Vazquez-Abad M-D, Shaw AR, Cappelleri JC, Ting N, Buckland-Wright C and Lustig S. Relationship between changes in joint space width (JSW) and C-reactive protein (CRP), clinical assessments, and health-related quality of life (HRQoL) instruments in mild/moderate osteoarthritis (OA). Arthritis and Rheumatism. 2000;43(9)(Suppl):S222.*

1041. *Wright JG and Young NL. The Patient-Specific Index: Asking patients what they want. Journal of Bone and Joint Surgery (Am). 1997;79(7):974-983.*

Guidance Documents for Outcome Measurement (WOMAC).

1042. *Abadie E, Ethgen D, Avouac B, Bouvenot G, Branco J, Bruyere O, Calvo G, Devogelaer J-P, Dreiser RL, Herrero-Beaumont G, Kahan A, Kreutz G, Laslop A, Lemmel EM, Nuki G, Van De Putte L, Vanhaels L, Reginstor J-Y, On Behalf of the Group for the Respect of Excellence and Ethics in Science (GREES). Recommendations for the use of new methods to assess the efficacy of disease-modifying drugs in the treatment of osteoarthritis. Osteoarthritis and Cartilage. 2004;12:263-268.*

1043. *Altman R, Brandt K, Hochberg M, Moskowitz R and the Task Force of the Osteoarthritis Research Society. Design and conduct of clinical trials in patients with osteoarthritis: Recommendations from a task force of the Osteoarthritis Research Society. Osteoarthritis and Cartilage. 1996;4:217-243.*

1044. *Bellamy N, Carr A, Dougados M, Shea B and Wells G. Towards a definition of "difference" in osteoarthritis. The Journal of Rheumatology. 2001;28:427-430.*

1045. *Bellamy N, Kirwan J, Boers M, et al. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. The Journal of Rheumatology. 1997;24:799-802.*

1046. *Bellamy N. Clinical trials design: structure modifying agents for osteoarthritis. Future guidelines: areas for development. Osteoarthritis and Cartilage. 1999;7:424-426.*

1047. *Bellamy N. Osteoarthritis clinical trials: Candidate variables and clinimetric properties. The Journal of Rheumatology. 1997;24:768-778.*

1048. *Bradley JD. Osteoarthritis treatment approaches, research methodologies, and agenda: do we need to re-invent the wheel? Arthritis Care and Research. 2000;13(5):252-254.*

1049. *Clinch J, Tugwell P, Wells G and Shea B. Individualized functional priority approach to the assessment of health related quality of life in rheumatology. The Journal of Rheumatology. 2001;28:445-451.*

1050. *Committee for Proprietary Medicinal Products (CPMP). Points to consider on clinical investigation of medicinal products used in the treatment of osteoarthritis. The European Agency for the Evaluation of Medicinal Products. London, 23 July 1998. EMEA http://www.eudra.org/emea.*

1051. *Dieppe PA. Recommended methodology for assessing the progression of osteoarthritis of the hip and knee joints. Osteoarthritis and Cartilage. 1995;3:73-77.*

1052. *Doherty M and Dougados M. EULAR recommendations for the treatment of knee OA. Osteoarthritis and Cartilage. 2000;8(Suppl B):S17.*

1053. *Doherty M and Jones A. Design of clinical trials in knee osteoarthritis: practical issues for debate. Osteoarthritis and Cartilage. 1998;6:371-373.*

1054. *Dougados M, Altman R, van der Heijde D, Le Claire P, Bloch D and Bellamy N. Preliminary OARSI responder criteria in knee and hip osteoarthritis. Annals of the Rheumatic Diseases. 2000;59(Suppl 1).*

1055. *Dougados M, Le Claire P, van der Heijde D, Bloch D, Bellamy N and Altman R. Preliminary definition of clinical response for osteoarthritis (OA) trials in knee and hip: a preliminary report of the Osteoarthritis Research Society International (OARSI) Standing Committee for clinical trials response criteria initiative. Arthritis and Rheumatism. 2000;43(9)(Suppl):S220.*

1056. Dougados M, LeClaire P, van der Heijde D, Bloch DA, Bellamy N and Altman RD. Special article: Response criteria for clinical trials on osteoarthritis of the knee and hip. A report of the Osteoarthritis Research Society International Standing Committee for clinical trials response criteria initiative. Osteoarthritis and Cartilage. 2000;8(6):395-403.

1057. Dougados M. Monitoring osteoarthritis progression and therapy. Osteoarthritis and Cartilage. 2000;8(Suppl B):S15.

1058. Dworkin R, Turk D, Farrar J, Haythornthwaite J, Jensen M, Katz N, Kerns R, Stucki G, Allen R, Bellamy N, Carr D, Chandler J, Cowan P, Dionne R, Galer B, Hertz S, Jadad A, Kramer L, Manning D, Martin S, McCormick C, McDermott M, McGrath P, Quessy S,Rappaport B, Robbins W, Robinson J, Rothman M, Royal M, Simon L, Stauffer J, Stein W, Tollett J, Wernicke J, and Witter J. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain 2004 (In Press).

1059. GREES (Group for the Respect of Ethics and Excellence in Science). Recommendations for the registration of drugs used in the treatment of osteoarthritis. Annals of the Rheumatic Diseases. 1996;55:552-557.

1060. Guidance for industry. Clinical development programs for drugs, devices, and biological products intended for the treatment of osteoarthritis (OA). Draft guidance. U.S. Department of Health and Human Services, FDA, July 1999.www.fda.gov/cder.

1061. Hammad TA. Structure modification in knee osteoarthritis: methodological outcome parameters. Osteoarthritis and Cartilage. 2001;9:488-498.

1062. Hochberg MC, Altman RD, Brandt KD and Moskowitz RW. Design and conduct of clinical trials in osteoarthritis: Preliminary recommendations from a task force of the Osteoarthritis Research Society. The Journal of Rheumatology. 1997;24:792-794.

1063. Jawad ASM. EULAR recommendations for the management of knee osteoarthritis. Annals of the Rheumatic Diseases. 2001;60:540.

1064. Jordan KM, Arden NK, Doherty M, Bannworth B, Bijlsma JWJ, Dieppe P, Gunther K, Hauelmann H, Herrero-Beaumont G, Kaklamanis P, Lohmander S, Leeb B, Lequesne M, Mazieres B, Martin-Mola E, Pavelka K, Pendleton A, Punzi L, Semi U, Swoboda B, Verbruggen G, Zimmerman-Gorska I and Dougados M. EULAR recommendations 2003: an evidence based approach to the management of knee osteoarthritis: Report of a Task Force of the Standing Committee for International Clinical Studies Including Therapeutic Trials (ESCISIT). Annals of the Rheumatic Diseases. 2003;62:1145-1155.

1065. Leeb BF. Management of knee osteoarthritis. Annals of the Rheumatic Diseases. 2001;60:984.

1066. Lequesne M, Brandt K, Bellamy N, Moskowitz R, Menkes CJ and Pelletier J-P. Guidelines for testing slow acting drugs in osteoarthritis. The Journal of Rheumatology. 1994;21(Suppl 41):65-73.

1067. Mazières B, Bannwarth B, Dougados M and Lequesne M. EULAR recommendations for the management of knee osteoarthritis. Report of a task force of the Standing Committee for International Clinical Studies Including Therapeutic Trials. Joint Bone Spine. 2001;68:231-240.

1068. Pendleton A, Arden N, Dougados M, et al. EULAR recommendations for the management of knee osteoarthritis: report of a task force of the Standing Committee for International Clinical Studies Including Therapeutic Trials (ESCISIT). Annals of the Rheumatic Diseases. 2000;59:936-944.

1069. Pham T, van der Heijde D, Lassere M, Altman RD, Anderson JJ, Bellamy N, Hochberg M, Simon L, Strand V, Woodworth T and Dougados M. Outcome variables for osteoarthritis clinical trials: the OMERACT-OARSI set of responder criteria. The Journal of Rheumatology. 2003;30(7):1648-1654.

1070. Strand V and Kelman A. Outcome measures in osteoarthritis: randomized controlled trials. Current Rheumatology Reports. 2004;6:20-30.

1071. Theiler R, Ghosh P and Brooks P. Clinical, biochemical and imaging methods of assessing osteoarthritis and clinical trials with agents claiming "chondromodulating" activity. Osteoarthritis and Cartilage. 1994;2:1-23.

1072. Turk DC, Dworkin RH, Allen RR, Bellamy N et al. Core outcome domains for chronic pain clinical trials: IMMPACT Recommendations. Pain 2003;106(3):337-45.

1073. Vignon E, Gamero P, Delmas P, Avouac B, Bettica P, Boers M, Ehrich E, Mackillop N, Rovati L, Sem U, Spector T, Reginster JY for the Group for the Respect of Ethics and Excellence in Science (GREES): Osteoarthritis Section. Recommendations for the registration of drugs used in the treatment of osteoarthritis: an update on biochemical markers. Osteoarthritis and Cartilage. 2001;9:289-293.

Miscellaneous

1074. Bellamy, N., Campbell, J. and Syrotuik, J.

1075. Comparative study of self-rating pain scales in osteoarthritis patients. Current Medical Research and Opinion 1999: 15(2),113-119.

1076. Bellamy, N., Campbell, J. and Syrotuik, J. Comparative study of self-rating pain scales in rheumatoid arthritis patients. Current Medical Research and Opinion 1999:15(2),121-127.

1077. Bellamy N. AUSCAN User Guide II, Brisbane, Australia 2003.

1078. Hobby KJ. The effects of aerobic walk-based exercise on women with rheumatoid arthritis. MSc Thesis, University of Western Ontario, London, Canada, 1994.

APPENDIX I - WOMAC 3.1 Alternate-Language Translations

LANGUAGE	3.1 LK	3.1/3.0 VA	3.1 NRS	3.01 LK	3.1 W	3.1 (IK)
Argentina	x	x				
Australia	x	x	x			
Austria	x	x			x	
Belgium French	x	x	x		x	
Belgium Flemish	x	x	x		x	
Brazil	x	x				
Brazil Japanese		x				
Bulgaria		x				
Canada English	x	x				x
Canada French	x	x				x
Chile	x	x				
China Mandarin		x				
Columbia	x	x				
Costa Rica		x				
Croatia	x	x			x	
Czech	x	x				
Denmark	x	x				
Ecuador	x					
Egypt Arabic		x				
Estonia		x				
Finland	x	x				
France	x	x	x		x	x
Germany	x	x	x		x	x
Greece		x				
Guatemala		x				
Hong Kong		x				
Hungary	x	x				
Iceland		x				
Israel	x	x				
Italy	x	x	x			
Japan	x	x				
Korea				x		
Latvia		x				

Lebanon			x			
Lithuania		x				
Malaysia Cantonese		x				
Malaysia English		x				
Malaysia Malay		x				
Mexico	x	x				
The Netherlands	x	x	x		x	
New Zealand	x	x				
Norway	x	x				
Peru	x	x				
Peru Japanese		x				
Philippines (Tagalog)		x				
Poland	x	x			x	
Portugal	x	x			x	
Romania		x				
Russian		x				
Singapore English		x				
Singapore Mandarin Chinese		x				
Slovak	x	x				
Slovenia	x	x				
South Africa (English)	x	x				
South Africa (Afrikaans)	x	x				
Spain	x	x	x		x	
Sweden	x	x				
Swiss French	x					
Swiss German	x	x				x
Swiss Italian	x					
Taiwan Mandarin Chinese		x				
Thailand		x				
Turkey	x	x				
United Kingdom	x	x			x	x
USA English	x	x	x			x
USA Spanish	x	x				
USA Florida		x				
Venezuela	x	x				

APPENDIX II - An Explanation Of The Meaning of Questions In The WOMAC Osteoarthritis Index Inventory

PAIN

Question 1:

"Walking on flat surface" refers to pain experienced while walking on even rather than uneven ground (i.e., walking in a shopping mall or on the sidewalk, or some other surface where there is a fair degree of regularity). This question does not refer to walking on uneven (i.e., rough) ground.

Question 2:

"Going up or down stairs" is self explanatory. If the pain is different going in one direction than the other, patients should rate according to the direction which produces the greatest pain.

Question 3:

"At night while in bed" refers to the kind of pain that disturbs sleep rather than that which occurs while lying in bed between going to bed and finally falling asleep, or between waking up and finally getting out of bed.

Question 4:

"Sitting or lying" refers to pain experienced either while in a position of sitting (i.e., in a chair) or while lying awake in bed.

Question 5:

"Standing upright" refers to pain occurring while in the standing position but not moving (cf Question 1).

STIFFNESS

Question 6:

Refers to the severity (rather than the duration) of stiffness which occurs after first awakening in the morning. In osteoarthritis this is usually, but not always, of short duration and often improves or disappears shortly after arising.

Question 7:

This question refers to the severity (rather than the duration) of stiffness which occurs after periods of inactivity later in the day. This is termed "gelling" in the literature.

These two questions have been phrased in this way because some patients seem to have a lot of morning stiffness but no gelling, and others have gelling with very little morning stiffness. Still other patients have both or neither.

PHYSICAL FUNCTION

Question 8:

Refers to the degree of difficulty descending stairs (irrespective of length, height or number).

Question 9:

Refers to the degree of difficulty ascending stairs (irrespective of length, height or number).

Question 10:

Refers to the degree of difficulty getting out of a chair (i.e., rising from the sitting position).

Question 11:

Refers to the degree of difficulty in remaining in a standing position and should not be confused with Question 13, which includes a dynamic component. It should also not be confused with Question 10 (i.e., it is the act of being in the standing position **not** the act of getting from **another** position **to** the standing position).

Question 12:

Refers to the degree of difficulty bending to pick something up off the floor. This usually involves some ankle movement, flexion of the knee and hip, and also some lumbar spinal flexion. Some patients seem to use more lumbar flexion and relatively little knee flexion, others appear to squat to pick up objects from the floor.

Question 13:

Refers to the degree of difficulty walking on a flat surface, that is an even surface such as a sidewalk or the inside of a shopping mall. It does not refer to walking on uneven (i.e., rough) ground.

Question 14:

Refers to the degree of difficulty getting in and out of a car, irrespective of whether this is into the driver's seat or a passenger seat. If the degree of difficulty differs between getting in versus getting out of a car, then the patient should rate the direction which produces the greatest difficulty. In later versions of the questionnaire, for use in a global context, this question may also encompass getting on and off a bus or other forms of transportation, since in some countries other forms of travel are more usual.

Question 15:

Refers to a composite activity which involves leaving a place of residence and negotiating the various obstacles and musculoskeletal challenges in the act of going shopping. This may include such simple impediments as getting on or off a curb, going up a slight rise, walking and standing for prolonged periods, and, in addition, is probably modulated by various social and emotional factors.

Question 16:

Refers to the degree of difficulty experienced while putting on socks or stockings. This question has been phrased to allow both male and female patients to respond.

Question 17:

Refers to the degree of difficulty getting out of bed (i.e., the act of swinging one's legs over the side and then getting into the standing position). This question differs from Question 10, in that the movement is made from a bed rather than a chair.

Question 18:

Refers to the degree of difficulty experienced while taking off socks and/or stockings and again has been modified so that male and female patients can both respond to the question.

Question 19:

Refers to the degree of difficulty lying in bed (i.e., turning from side to side, or maintaining one particular position in the lying posture).

Question 20:

Refers to the degree of difficulty in getting in and out of the bath tub. For patients who take a shower, this question could refer to the shower rather than the bath. If the difficulty differs between getting in and out of the bath, then the patient should rate that activity which produces the greatest difficulty.

Question 21:

Refers to the degree of difficultly being in a sitting position, i.e., static positioning or shuffling about in a chair during prolonged sitting. This is in contra-distinction to Question 10, which asks about rising from the sitting position.

Question 22:

Refers to getting on or off the toilet. If the degree of difficulty is different for the two actions, then the patient should rate that action which produces the most difficulty. In countries where toilets are non-Western in style, the question could pertain to the relevant biomechanical challenges relating to this activity.

Question 23:

Refers to heavy domestic duties. It has been phrased in these terms to allow both male and female patients to respond. Heavy domestic duties for a male might include mowing the lawn, raking leaves, shovelling snow or moving heavy boxes, etc. Heavy domestic duties for a female might include vacuuming, moving heavy boxes, scrubbing floors, lifting heavy grocery bags, etc. There are, of course, various other examples.

Question 24:

Refers to light domestic duties. Again this has been phrased to allow both male and female patients to respond. Light domestic duties for a male might include tidying up a room, indulging in crafts or hobbies, laying or clearing a table, etc. Light domestic duties for a female might include cooking a meal, laying and clearing a table, dusting, indulging in crafts and hobbies etc. There are, of course, various other examples.

AUSCAN HAND OSTEOARTHRITIS INDEX

The AUSCAN Hand Osteoarthritis Index is a tridimensional, self-administered, patient-centered health status questionnaire. Its item inventory has been designed to capture the essential elements of pain, stiffness and physical disability in patients with osteoarthritis of the hand joints. The AUSCAN Index has been subject to studies examining its basic clinimetric properties of reliability, validity and responsiveness and the AUSCAN Index has been translated into over twenty different language forms. It is available in 5-point Likert (LK), 100 mm Visual Analogue (VA) and 11-point Numerical Rating Scale (NRS) formats.

LANGUAGE	3.1 LK	3.0 VA	3.1 NRS
Australia	x		x
Austria	x		
Belgium French			x
Belgium Flemish			x
Canada English	x	x	
Canada French	x	x	
Czech	x		
France	x		x
Finland	x		
Germany	x		x
Hungary	x		
Israel	x		
Italy	x		x
Lebanon			x
The Netherlands	x		x
New Zealand	x		
Norway	x		
Poland	x		
Russia	x		
Slovakia	x		
South Africa (English)	x		
South Africa (Afrikaans)	x		
Spain	x		x
Sweden	x		
Turkey	x		
United Kingdom			x
USA English	x		
USA Spanish		x	

AUSCAN website: www.auscan.org

OSTEOARTHRITIS GLOBAL INDEX (OGI)

The OGI 8.0 is a self-administered questionnaire that assesses the beneficial effects of therapy at three different levels (study joint, disease, person) and grades adverse effects of therapy, using a battery of 8 questions.

OGI 8.0 uses Likert-type scaling, and has been used in whole or in part in two recent Canadian studies.

Bellamy N, Goldstein LD and Tekanoff RA. Continuing medical education-driven skills acquisition and impact on improved patient outcomes in family practice setting. *The Journal of Continuing Education in the Health Professions*;Vol 20,(1), Winter 2000: pp52-61.

Jean Pierre Raynauld, George W Torrance, Philip A Band, Charles H Goldsmith, Peter Tugwell, Valery Walker, Margarita Schultz, and Nicholas Bellamy, in collaboration with the Canadian Knee OA Study Group. *Osteoarthritis & Cartilage* 2002:10(7);517-526.

The OGI 8.0 is an initial step in the development of a standard measurement battery for performing patient global assessments. The OGI 8.0 is currently available in English for North America and French for Canada.

OGI website: www.ogiq.org

SECTION 10 – WOMAC Index: Contemporary Context

Twenty-three years have passed since the original conceptualization of the WOMAC Index, and seventeen years since the first WOMAC Index validation studies were published. From a rheumatology perspective, the emergence of new measurement techniques in the 1980s, the recommendations of the OMERACT and IMMPACT groups, the need for regulatory agencies such as the FDA and EMEA to establish guidelines and the publication of guidelines for the conduct of clinical trials by societies such as the OARSI, has generated increased interest in the field of outcome measurement in OA. It is important to understand the WOMAC Index, from the standpoint of its contemporary context, that is, its relationship to other indices of osteoarthritis clinical severity and impact, particularly with respect to measurement tools developed since publication of the WOMAC validation studies in 1988. Although this thesis primarily concerns use of the WOMAC Index in rheumatology environments, it is noteworthy that the emphasis on certain aspects of measurement, may vary between the health disciplines of rheumatology, physical therapy, orthopaedic surgery and rehabilitation medicine. For example, while not exclusive to any one discipline, the biomechanical aspects of the condition are particularly important in orthopaedic environments, while functional independence is especially important in physical therapy and rehabilitation environments. As a consequence, a large number of evaluation techniques have evolved to measure the impact of various conditions on different health, social, emotional, vocational and economic outcomes. In this section, the contemporary context will be considered from the perspective of questionnaire-based patient-centred symptom severity evaluation tools for hip and knee OA, that have been developed, since the publication in 1988 of the WOMAC validation studies, and which are finding application in rheumatology.

The contemporary measurement environment in OA hip and knee assessment in rheumatology can be divided according to the continuing evolution of experience with patient-centred measurement techniques whose origins predate the WOMAC Index, and those entirely new assessment techniques which have emerged since the WOMAC was validated. Conceptually the measurement framework in hip and knee OA, includes three types of patient-centred measures: Generic HRQOL, General Arthritis Measures and OA-Specific Measures (Appendix B).

Of the generic HRQOL measures, the SF-36, Nottingham Health Profile (NHP), European Quality of Life (EuroQoL), and the Health Utilities Index (HUI), had already found application in OA prior to the development of the WOMAC Index. The use of these measures, their derivatives such as SF-12 and HUI-3, and alternate-language translations, continues to expand opportunities for evaluating HRQOL in OA, from a generic perspective. Furthermore, in 1999, Ware and colleagues described a methodology for developing an Arthritis-Specific Health Index (ASHI) score form the SF-36. More recent additions to the list of generic HRQOL instruments, include the World Health Organisation Quality of Life Measure (WHO-QoL). Experience with this measure is growing, and its exact position within the hierarchy of generic HRQOL tools, and its use in OA, remain to be established. Multinational interest in this measure, suggests that it will be of increasing importance, particularly in the public health sector.

Of the general arthritis measures the HAQ and AIMS instruments, and their derivatives such as MHAQ and AIMS2, have remained in common usage, where a broad based measurement of OA status is required. Their position in OA measurement has not been seriously challenged in the last 25 years, and they are useful, where attribution to a single joint, of the severity and consequence of OA, in patients who have differing patterns of multi-joint OA involvement, is not required. The development of variations, short forms and multiple alternate-language translations continues, and highlight the importance of this group of measures.

The WOMAC Index and the ICS remain the two most commonly used hip and knee patient-centred OA-specific outcome measures reported in the rheumatology interventional research literature. Nevertheless, several new measures have emerged in the last seventeen years, have been validated in OA patients, and are finding application in rheumatology (Appendix B).

Brooks and colleagues in Australia have developed an approach to outcome measurement in OA, based on reducing respondent burden, by including only four patient self-reported global questions, in a measure termed the Comprehensive Osteoarthritis Test (COAT) (Appendix B). The construct validity of the COAT has been established using the WOMAC Index. The COAT can be completed quickly and scored easily. While methods based on patient global impression have merit, the exact process by which the patient reviews their symptom experience, and then selects, weights and aggregates the information into a global score, is poorly understood in OA patients. Further research on the nature of patient global assessment, and the exact wording of patient global questions, has been encouraged by the OARSI Task Force on Guidelines for Clinical Trials (23). Given its recent introduction, it is too early to comment on the uptake of the COAT in clinical practice or clinical research.

The Joint-Specific Multidimensional Assessment of Pain (J-MAP), is a pain measurement tool containing both sensory and affective items (Appendix B). The 6-item Pain Sensory and 4-item Pain Affect subscales, are patient self-completed. The measure is valid, reliable and responsive, and assesses pain from a joint-specific perspective.

The Knee Injury and Osteoarthritis Outcome Score (KOOS) and the Hip Disability and Osteoarthritis Outcome Score (HOOS), have both been developed in Sweden (Appendix B). These two patient self-completed questionnaires, attempt to encompass the needs of a broader spectrum of patients with OA. Parts of the KOOS and HOOS Indices borrow content from the WOMAC Index. The WOMAC Index was used by Roos and colleagues, in establishing the content validity of the KOOS Index, while Nilsdotter et al used the WOMAC to establish the content validity of the HOOS Index. The KOOS and HOOS Indices are valid, reliable and responsive. The KOOS Index is being used with increasing frequency in clinical research and clinical practice, more so in orthopaedic environments than in rheumatology. To date, uptake of the HOOS appears to have been limited, although this may, in part, relate to its relatively recent emergence.

The Oxford Knee Score (OKS), and the Oxford Hip Score (OHS), are also valid, reliable and responsive measures of outcome, and have found application, particularly in orthopaedic environments, to evaluate the impact of total joint replacement surgery (Appendix B). Both scales, include 12-items and are patient self-completed. In general, these two instruments have performed well, although Whitehouse and colleagues have

recently questioned, whether the formulation of OKS questions and response categories needs reconsideration.

The Short Arthritis Assessment Scale (SAS), has been developed from existing measures and includes two WOMAC items (descending stairs and going shopping), together with one pain and one patient global item (Appendix B). The measure is brief, simple to score, and may find application in routine clinical practice. However, the physical function component of the SAS, lacks the content validity of the physical function subscale of the WOMAC Index, from which it is derived (2 items vs 17 items).

The German Short Musculoskeletal Function Assessment Questionnaire (SMFA-D), contains two subscales (function index and bother index), and includes 46 items (Appendix B). The WOMAC Index has been used to establish the construct validity of the SMFA-D Index. The SMFA-D is a valid, reliable and responsive measure of outcome in OA patients.

The contemporary context within which the WOMAC Index exists, is characterized by a) formalized measurement guidelines based on evidence and the consensus of expert opinion, b) requirements for assessment tools that are valid, reliable and responsive, c) different classes of tools that meet different measurement requirements (generic HRQOL, general arthritis and condition specific), and d) within each class a choice of tools which differ in their concept, content, application and performance. From a conceptual standpoint, the condition specific instruments, like the WOMAC Index, vary in the extent to which patient opinion has been incorporated in the index development, and whether the instrument generates subscale scores or a composite score. Their content varies according to the number and nature of the dimensions covered by the question inventory. With respect to application, the instruments vary in their mode of administration (patient self-completed or interviewer administered), mode of delivery (paper, telephone, electronic), time frame, scaling format, and the availability of alternate-language translations. The clinimetric performance of the various instruments differs. While their reliability is generally high, they differ in their sensitivity to change (syn: responsiveness), and their validity. Thus, some instruments are more valid for certain applications. The existence of measures that differ in concept and content, creates measurement opportunities, particularly when multiple measures are used in combination. For example the combined use of the WOMAC Index, SF-36 and HUI-3 in the studies reported by Raynauld et al (17) and Torrance et al (18), permitted the respective issues of effectiveness, cost-effectiveness and cost-utility to be comprehensively addressed.

The WOMAC Index has not only contributed to meeting the measurement needs of significant numbers of researchers and clinicians, but it has also played an important role in the development and validation of new health status measures. A robust clinimetric profile (validity, reliability, responsiveness), flexible delivery, multiple scaling options and a large inventory of linguistically valid alternate-language translations, as well as a central role in setting responder and state-attainment criteria, has firmly established the relevance and importance of the WOMAC Index in a conventional context, and encompassing diverse and evolutionary measurement needs.

REFERENCES

1. Bellamy, N. and Buchanan, W.W. A preliminary evaluation of the dimensionality and clinical importance of pain and disability in osteoarthritis of the hip and knee. Clinical Rheumatology 1986;5(2):231-241.

2. Bellamy, N., Buchanan, W.W., Goldsmith, C.H., Campbell, J. and Stitt, L. Validation study of WOMAC: a health status instrument for measuring clinically-important patient-relevant outcomes following total hip or knee arthroplasty in osteoarthritis. Journal of Orthopaedic Rheumatology 1988;1:95-108.

3. Bellamy, N., Buchanan, W.W. and Grace, E. Double-blind randomized controlled trial of isoxicam vs piroxicam in elderly patients with osteoarthritis of the hip and knee. British Journal of Clinical Pharmacology 1986;22:149S-155S.

4. Bellamy, N., Buchanan, W.W., Goldsmith, C.H., Campbell, J. and Stitt, L. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. The Journal of Rheumatology 1988;15(12):1833-1840.

5. Bellamy, N., Goldsmith, C.H., Buchanan, W.W., Campbell, J. and Duku, E. Prior score availability: Observations using the WOMAC Osteoarthritis Index. British Journal of Rheumatology 1991;30:150-151.

6. Bellamy, N., Wells, G. and Campbell, J. Relationship between severity and clinical importance of symptoms in osteoarthritis. Clinical Rheumatology 1991;10(2):138-143.

7. Bellamy, N., Buchanan, W.W., Goldsmith, C.H., Campbell, J. and Duku, E. Signal measurement strategies: are they feasible and do they offer any advantage in outcome measurement in osteoarthritis? Arthritis and Rheumatism 1990;33(5):739-745.

8. Barr, S., Bellamy, N., Buchanan, W.W., Chalmers, A., Ford, P.M., Kean, W.F., Kraag, G.R., Gerecz-Simon, E. and Campbell, J. A comparative study of signal versus aggregate methods of outcome measurement based on the WOMAC Osteoarthritis Index. The Journal of Rheumatology 1994;21(11):2106-2112.

9. Griffiths, G., Bellamy, N., Kean, W.F., Campbell, J. and Gerecz-Simon, E. A study of the time frame dependency of responses to the WOMAC Osteoarthritis Index. Inflammopharmacology 1993;2:85-87.

10. Griffiths, G., Bellamy, N., Bailey, W.H., Bailey, S.I., McLaren, A.C. and Campbell, J. A comparative study of the relative efficiency of the WOMAC, AIMS and HAQ instruments in evaluating the outcome of total knee arthroplasty. Inflammopharmacology 1995;3:1-6.

11. Davies, G.M., Watson, D.J., and Bellamy, N. Comparison of the responsiveness and relative effect size of the Western Ontario and McMaster Universities Osteoarthritis Index and the Short-Form Medical Outcomes Study Survey in a randomized, clinical trial of osteoarthritis patients. Arthritis Care and Research 1999;12(3):172-179.

12. Bellamy, N., Kean, W.F., Buchanan, W.W., Gerecz-Simon, E. and Campbell, J. Double blind randomized controlled trial of sodium meclofenamate (Meclomen) and diclofenac sodium (Voltaren): Post validation reapplication of the WOMAC Osteoarthritis Index. The Journal of Rheumatology 1992;19(1):153-159.

13. Bellamy, N., Buchanan, W.W., Chalmers, A., Ford, P.M., Kean, W.F., Kraag, G.K., Gerecz-Simon, E. and Campbell, J. A multicenter study of tenoxicam and diclofenac in patients with osteoarthritis of the knee. The Journal of Rheumatology 1993;20(6):999-1004.

14. Peloso, P.M., Bellamy, N., Bensen, W., Thomson, G.T.D., Harsanyi, Z., Babul, N., and Darke, A.C. Double blind randomized placebo control trial of controlled release codeine in the treatment of osteoarthritis of the hip or knee. Journal of Rheumatology 2000;27(3):764-771.

15. Pelletier, J-P., Yaron, M., Haraoui, B., Cohen, P., Menahem A. Nahir, M.A., Choquette, D., Wigler, I., Rosner, I.A., Beaulieu, A.D. and The Diacerein Study Group. Efficacy and safety of diacerein in osteoarthritis of the knee. Arthritis & Rheumatism 2000;43(10):2339-2348.

16. Bellamy, N., Goldstein. L.D. and Tekanoff, R.A. Continuing Medical Education-driven skills acquisition and impact on improved patient outcomes in family practice setting. The Journal of Continuing Education in the Health Professions 2000;20:52-61.

17. Raynauld, J.P., Torrance, G.W., Band, P.A., Goldsmith, C.H., Tugwell, P., Walker, V., Schultz, M., Bellamy, N. in collaboration with the Canadian Knee OA Study Group. A prospective, randomized, pragmatic, health outcomes trial evaluating the incorporation of hylan G-F 20 into the treatment paradigm for patients with knee osteoarthritis (Part 1 of 2): clinical results. Osteoarthritis and Cartilage 2002;10(7):506-517.

18. Torrance, G.W., Raynauld, J.P., Walker, V., Goldsmith, C.H., Bellamy, N., Band, P.A., Schultz, M. and Tugwell, P. A prospective, randomized, pragmatic, health outcomes trial evaluating the incorporation of hylan G-F 20 into the treatment paradigm for patients with knee osteoarthritis (Part 2 of 2): economic results. Osteoarthritis and Cartilage 2002;10(7):518-527.

19. Lequesne, M., Brandt, K., Bellamy, N., Moskowitz, R., Menkes, C.J., Pelletier, J-P. and Altman, R. Guidelines for testing slow acting drugs in osteoarthritis. The Journal of Rheumatology 1994;(Suppl 41)21:65-73.

20. Bellamy, N. Outcome measurement in osteoarthritis clinical trials. The Journal of Rheumatology 1995;(Suppl 43)22:49-51.

21. Bellamy, N. Osteoarthritis clinical trials: Candidate variables and clinimetric properties. The Journal of Rheumatology 1997;24(4):768-778.

22. Bellamy, N., Kirwan, J., Boers, M., Brooks, P., Strand, V., Tugwell, P., Altman, R., Brandt, K., Dougados, M. and Lequesne, M. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip and hand osteoarthritis. Consensus development in OMERACT III. The Journal of Rheumatology 1997;24(4):799-802.

23. Osteoarthritis Research Society (OARS) Task Force Report: Design and conduct of clinical trials in patients with osteoarthritis: Recommendations from a Task Force of the Osteoarthritis Research Society. Osteoarthritis and Cartilage 1996;4(4):217-243.

24. Turk, D.C., Dworkin, R.H., Allen, R.R., Bellamy, N. et al. Core outcome domains for chronic pain clinical trials: IMMPACT recommendations. Pain 2003;106(3):337-345.

25. Dworkin, R.H., Turk, D.C., Farrar, J.T., Haythornthwaite, J.A., Jensen, M.P., Katz, N.P., Kerns, R.D., Stucki, G., Allen, R.R., Bellamy, N. et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain 2005;113:9-19.

26. Bellamy N. WOMAC: A 20-year experiential review of a patient-centered self-reported health status questionnaire. The Journal of Rheumatology 2002;29(12):2473-2476.

27. Bellamy, N., Campbell, J., Stevens, J., Pilch, L., Stewart, C. and Mahmood, Z. Validation study of a computerized version of the Western Ontario and McMaster Universities VA3.0 Osteoarthritis Index. The Journal of Rheumatology 1997;24(12):2413-2415.

28. Bellamy,N., Campbell,J., Hill, J. and Band, P. A comparative study of telephone versus onsite completion of the WOMAC 3.0 Osteoarthritis Index. The Journal of Rheumatology 2002;29(4):783-786.

29. Theiler, R., Spielberger, J., Bischoff, H.A., Bellamy, N., Huber, J. and Kroesen, S. Clinical evaluation of the WOMAC 3.0 OA Index in numeric rating scale format using a computerized touch screen version. Osteoarthritis and Cartilage 2002;10(6):479-481.

30. Theiler, R., Bischoff-Ferrari, H.A., Good, M. and Bellamy, N. Responsiveness of the electronic touch screen WOMAC 3.1 OA Index in a short term clinical trial with rofecoxib. Osteoarthritis and Cartilage 2004;12(11):912-916.

31.Bellamy, N., Campbell, J. and Syrotuik, J. Comparative study of self-rating pain scales in osteoarthritis patients. Current Medical Research and Opinion 1999;15(2):113-119.

32. Bellamy, N., Kaloni, S., Pope, J., Coulter, K. and Campbell, J. Quantitative rheumatology: A survey of outcome measurement procedures in routine rheumatology outpatient practice in Canada. The Journal of Rheumatology 1998;25(5):852-858.

33. Bellamy, N., Muirden, K.D., Brooks, P.M., Barraclough, D., Tellus, M.M. and Campbell, J. A survey of outcome measurement procedures in routine rheumatology outpatient practice in Australia. The Journal of Rheumatology 1999;26(7):1593-1599.

34. Tubach, F., Baron, G., Falissard, B., Logeart, I., Dougados, M., Bellamy, N. and Ravaud, P. Using patients' and rheumatologists' opinions to specify a short form of the WOMAC function subscale. Annals of the Rheumatic Diseases 2005;64(1):75-79.

35. Dougados, M., LeClaire, P., van der Heijde, D., Bloch, DA., Bellamy, N. and Altman, RD. Special article: Response criteria for clinical trials on osteoarthritis of the knee and hip: A report of the Osteoarthritis Research Society International Standing Committee for Clinical Trials Response Criteria Initiative. Osteoarthritis and Cartilage 2000;8(6):395-403.

36. Ehrich, E.W., Davies, G.M., Watson, D.J., Bolognese, J.A., Seidenberg, B.C. and Bellamy, N. Minimal Perceptible Clinical Improvement with the Western Ontario and McMaster Universities Osteoarthritis Index questionnaire and global assessments in patients with osteoarthritis. The Journal of Rheumatology 2000;27(11):2635-2641.

37. Bellamy, N., Carr, A., Dougados, M., Shea, B. and Wells, G. Towards a definition of "difference" in osteoarthritis. The Journal of Rheumatology 2001;28(2):427-430.

38. Wells, G., Anderson, J., Beaton, D., Bellamy, N., Boers, M., Bombardier, C., Breedveld, F., Carr, A., Cranney, A., Dougados, M., Felson, D., Kirwan, J., Schiff, M., Shea, B., Simon, L., Smolen, J., Strand, V., Tugwell, P., van Riel, P. and Welch, V.A. Minimal Clinically Important Difference Module: Summary, Recommendations, and Research Agenda. The Journal of Rheumatology 2001;28(2):452-454.

39. Pham, T., van der Heijde, D., Altman, R.D., Anderson, J.J., Bellamy, N., Hochberg, M., Simon, L., Strand, V., Woodworth, T. and Dougados, M. OMERACT-OARSI Initiative: Osteoarthritis Research Society International set of responder criteria for osteoarthritis clinical trials revisited. Osteoarthritis and Cartilage 2004;12(5):389-399.

40. Bellamy, N., Bell, M.J., Goldsmith, C.H., Pericak, D., Walker, V., Raynauld, J-P., Torrance, G., Tugwell, P. and Polisson, R. Evaluation of WOMAC 20, 50, 70 response criteria in patients treated with hylan G-F 20 for knee osteoarthritis. Annals of the Rheumatic Diseases 2005;64:881-885.

41. Bellamy, N., Bell, M.J., Goldsmith, C.H., Pericak, D., Walker, V., Raynauld, J-P., Torrance, G.W., Tugwell, P. and Polisson, R. The effectiveness of hylan G-F 20 in patients with knee osteoarthritis: an application of two sets of response criteria developed by the OARSI and one set developed by OMERACT-OARSI. Osteoarthritis and Cartilage 2005;13(2):104-110.

42. Tubach, F., Ravaud, P., Baron, G., Falissard, B., Logeart, I., Bellamy, N., Bombardier, C., Felson, D., Hochberg, M., van der Heijde, D. and Dougados, M. Evaluation of clinically relevant states in patient reported outcomes in knee and hip osteoarthritis: the minimal clinically important improvement. Annals of the Rheumatic Diseases 2005;64(1):29-33.

43. Tubach, F., Ravaud, P., Baron, G., Falissard, B., Logeart, I., Bellamy, N., Bombardier, C., Felson, D., Hochberg, M., van der Heijde, D. and Dougados, M. Evaluation of clinically relevant states in patient reported outcomes in knee and hip osteoarthritis: the patient acceptable symptom state. Annals of the Rheumatic Diseases 2005;64(1):34-37.

44. Bellamy, N. WOMAC Osteoarthritis Index – User Guide VII (Published by N. Bellamy, Brisbane, Australia) 2004.

APPENDIX B

OUTCOME MEASURES FOR HIP AND KNEE OA IN RHEUMATOLOGY: CONTEMPORARY CONTEXT

GENERIC HRQOL MEASURES:

European Quality of Life Index (EuroQoL) - Hurst NP et al. Br J Rheumatol. 1994;33:655-662.

Health Utilities Index (HUI) - Feeny D et al. J Clin Oncol. 1992;10:923-928.

Nottingham Health Profile (NHP) – Hunt et al. Soc Sci Med 1981;15A:221-229.

World Health Organisation Quality of Life Index (WHO-QoL)* – Hawthorne et al. Quality of Life Res 1999;8:209-224.

Short Form 36 (SF-36) - Ware JE et al. Med. Care. 1992;30:473-483.

GENERAL ARTHRITIS MEASURES:

Arthritis Impact Measurement Scales (AIMS) - Meenan RF et al Arthritis Rheum. 1980;23:146-152.

Arthritis Impact Measurement Scales 2 (AIMS2) - Meenan RF et al. Arthritis Rheum. 1992;35:1-10.

Health Assessment Questionnaire (HAQ) - Fries JF et al. Arthritis Rheum. 1980;23:137-145.

Modified Health Assessment Questionnaire (MHAQ) – Pincus T et al. Arthritis Rheum 1983;26:1346-1353.

CONDITION-SPECIFIC AND JOINT-SPECIFIC MEASURES:

Comprehensive Osteoarthritis Test (COAT)* - Brooks LO et al. J Rheumatol 2004;31:1180-1186.

Hip Disability and Osteoarthritis Outcome Score (HOOS)* – Nilsdotter AK et al. BMC Musc Disord 2003;4:10-17.

Index of Clinical Severity (Hip) - Lequesne MG et al. Scand J Rheumatol. 1987;65(Suppl):85-89.

Index of Clinical Severity (Knee) - Lequesne MG et al. Scand J Rheumatol. 1987;65(Suppl):85-89.

Joint-Specific Multidimensional Assessment of Pain (J-MAP)* - O'Malley et al. J Rheumatol 2003;30:534-543.

Knee Injury and Osteoarthritis Outcome Score (KOOS)* - Roos EM et al. J Orthop Sports Phys 1998;78:88-96.

Oxford Hip Score (OHS)* – Dawson J et al. J Bone Joint Surg, Br 1996;78-B:185-190.

Oxford Knee Score (OKS)* – Dawson J et al. J Bone Joint Surg, Br 1998;80-B:63-69.

Short Arthritis Assessment Scale* - Wolfe et al. J Rheumatol. 2004;31:2472-2479.

Short Musculoskeletal Function Assessment Questionnaire (SMFA-D)* - Kirschner S et al. Rheumatol Int 2003;23:15-20.

Western Ontario and McMaster Osteoarthritis Index (WOMAC) - Bellamy N et al. J Rheumatol. 1988;15:1833-1840.

[* Measures developed since publication of the WOMAC Index validation studies]