

**Measuring change in patient quality of life
over time: an evaluation of scale
responsiveness and patient response shift**

by

Michael Robling

UMI Number: U584076

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U584076

Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Summary

Measuring change in quality of life is increasingly central to health services and clinical research evaluation. This requires instruments that are responsive to change, and that the construct being assessed is stable. I have, therefore, addressed two methodological themes: scale responsiveness and instability of the underlying quality of life construct - response shift.

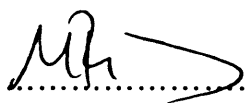
Responsiveness theme: I evaluated performance characteristics of a commonly reported effect size statistic, the standardised response mean (SRM). Computer simulations modelled the impact of varying computational method and distributional characteristics upon bias of estimated effect size compared to underlying true value. The studies provide evidence and reassurance that the SRM exhibits little bias when sample size, mean underlying effect size and shape of underlying distribution are varied. However, alternate approaches to handling negative values can produce markedly different effect sizes, making comparison across studies that use different methods problematic. Furthermore, parametric SRMs calculated from lognormal data may provide a greatly inflated estimate of effect size.

Response shift theme: I interviewed patients at different stages of clinical management for knee injury twice over six months. A multi-method approach incorporating the individualised SEIQoL-DW measure and a retrospective pretest-posttest using EQ-5D identified evidence of re-calibration, re-prioritisation and re-conceptualisation response shift. An empirically based typology of changes was developed drawn from existing response shift theory, but which further distinguishes subtler forms of change. The studies provide evidence that re-prioritisation and re-conceptualisation may be different levels of the same process. Furthermore, mechanisms producing response shift were identified, in particular, the interaction between level of satisfaction with quality of life domain and its perceived importance. Additional approaches to studying response shift using group level comparison of SEIQoL data were critically evaluated.

The thesis extends the methods for identifying, assessing and conceptualising response shift changes whilst also exploring mechanisms which may explain these changes.

DECLARATION

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed (candidate)

Date 19/10/06

STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated.

Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed (candidate)

Date 19/10/06

STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed (candidate)

Date 19/10/06

Acknowledgements

This thesis was based on an original conversation during a car journey between North and South Wales ... travel truly broadens the mind.

I am grateful to my initial and subsequent supervisors, Professor Paul Bennett and Professor Chris Butler, who in turn helped me get off the mark and finish the job. Paul, thanks in particular for encouraging me to carry on past the initial stages of my work and to consolidate my research schedule. I gratefully acknowledge my scientific and clinical collaborators who have helped make this thesis possible. Thus, I would like to thank Professor Ian Russell for his constructive methodological advice on the responsiveness modelling, and Dr Kerry Hood for her consistently accessible statistical expertise. I am also grateful to Ian and Dr Stephen Brealey, as Principal Investigator and Trial Co-ordinator respectively of the DAMASK Study, for allowing me to develop the response shift studies alongside that project. I would also like to thank the local clinical collaborators on The DAMASK Study, Mr Chris Wilson and Dr Kathleen Lyons for facilitating access to study patients. I am grateful to Professor Roisin Pill for her advice regarding the design and analysis of the response shift study. Our discussions about the qualitative methods have been a satisfying synthesis of my uncertainty and your re-assurance. Thank you also to Dr Fiona Wood for her advice on the analysis and presentation of the qualitative data.

Several people have transcribed the many interviews conducted in the course of this study. Thank you therefore to Jo Sloan, Sheila Morris, Penny Moore, Mandy Iles and Helen Evans. I am grateful to the Scientific Foundation Board of the Royal College of General Practitioners, who kindly funded the costs of interview conduct and transcription. I am also grateful to the head of the Department of General Practice, Professor Helen Houston, for her continued support for me during the conduct of this work. I would also like to acknowledge the enthusiasm and support of my many colleagues within my department.

I am indebted to Dr Lionel Jacobson & Mrs Rebecca Cannings-John for their generosity in offering to proof-read my thesis draft, and for their skill and precision in doing so. Becky, I hope that you offer me the opportunity of repaying you when you draft your own thesis. Lionel, thank you also for your wise reflections at various stages of my studies – the supportive spirit behind your comments as welcome as their content.

This thesis would not have been possible without the selfless participation of all the patients interviewed. The opportunity to meet with them and hear their stories has been a privilege. I hope that this thesis does justice to all those who have contributed to its conduct, content and presentation.

In conducting the work of this thesis, I have been fortunate to have worked with many friends and colleagues who have variously acted to guide, inspire, mentor and support me along the way. However, two people can tick all of

these boxes and more, Professor Chris Butler and Dr Kerry Hood. Chris, as my supervisor I particularly thank you for asking me the tough questions about my work and for helping me to keep an eye on the bigger picture. It is a wonderful skill to be able to provide quick, insightful and always constructive feedback and advice, and you have it in abundance. Kerry, thank you for your support with almost every facet of this work – from discussing initial ideas, statistical advice in both the responsiveness and response shift studies, continual encouragement and strategic advice at critical stages of development. It is my good fortune to have been able to work with two colleagues for whom I have such professional and personal respect.

Thank you Sue for your company, support, tolerance and sacrifice whilst I have undertaken this research. Without you this thesis would not have been possible.

Dedication: This thesis is dedicated to my late parents, Louis and Susan Robling, who continue to enrich the quality of my own life through my fondest memories.

Author's forward

Abstract: *This forward is a personal account of the development of my interest in response shift and scale responsiveness - the subject matter of my thesis. It describes how this interest was stimulated by unresolved methodological challenges encountered whilst working as a clinical trial researcher. These issues related to the assessment of change using health outcome measures. The forward outlines how I developed the aims and objectives of my thesis and what I sought to contribute to knowledge and theory of change assessment.*

Working as a clinical trialist: I was working as the research fellow on a pragmatic cluster randomised controlled trial in 1999, evaluating the impact of primary care guidelines for the management of women with breast problems (The BRIDGE Study).¹ Nearly forty group general practices across South Wales, and over 2000 patients, were recruited into the study during a baseline epidemiological phase and a subsequent intervention phase. The research team evaluated a complex intervention incorporating a guidelines dissemination and implementation package.² The research team aimed to assess the impact of the intervention upon the clinical management of patients and patient-related outcomes (including quality of life). Could the guidelines facilitate referral of patients where appropriate; enable primary care management of patients with benign problems without referral, whilst supporting the well-being of all patients?

Research evaluating such health service innovation demands the availability of suitable patient-based outcome measures. Outcome in the BRIDGE Study was assessed with a questionnaire which included established generic health status measures (the MOS SF-36 and EQ-5D) and a specific measure.^{3 4} Patients in the trial were being assessed six and 12 months following an index consultation with a study general practitioner.

Developing an outcome measure: The original study protocol indicated that a specific measure to assess outcome in women with benign breast disease would have to be developed if one did not already exist. An initial review confirmed that such scale development was required. Item identification and selection was based on qualitative interviews with primary and secondary care clinicians; literature review; pre-piloting; and cognitive debrief interviews with patients. Candidate items were included in an outcome questionnaire package alongside the generic measures indicated above. Questionnaires were disseminated, completed and the response rate was satisfactory. We analysed the factor structure of the potential scale items and eventually the three sub-scales of the Cardiff Breast Scales emerged.⁵

Throughout this process, standard procedures for scale development were followed.⁶ Particular attention was paid to assessing, and subsequently reporting, scale attributes of validity and reliability. As a graduate of applied psychology these were familiar psychometric concepts, but these are not the only considerations when developing and reporting a new instrument. A contemporary monograph from the NHS R&D Health Technology Assessment programme had outlined eight key criteria for the selection of outcome measures in trials.⁷ One such criteria listed was scale responsiveness – the ability to detect clinically important change over time.

Identifying a problem: In the context of randomised controlled trials, the importance of using responsive outcome measures appeared self-evident. An unresponsive measure would fail to detect important clinical change when present and could lead to erroneous conclusions about intervention efficacy. At the very least, more patients, money and time would be required to demonstrate the same effect compared to a more responsive measure. What surprised me was the lack of consensus about how this attribute should be assessed and reported. A variety of approaches were described but, as Fitzpatrick and colleagues stated, ‘the literature on responsiveness is not as well developed as it is for reliability and validity’.⁷ Even for some of the main responsiveness statistics there were a variety of ways by which they could be calculated. To complete the development and reporting of the new instrument,

we chose a common approach (the standardised response mean) but many questions were left unanswered.⁵

Developing my ideas: I decided that contributing to this area would be the topic of my thesis. In formulating ideas, I considered the possibility of comparing the main approaches to assessing responsiveness, utilising data accrued from the BRIDGE Study. Change scores, standardised effect size (SES), standardised response mean (SRM), Guyatt's Index of Responsiveness (IoR) and relative efficiency (RE) could be compared. The Cardiff Breast Scales and the generic measures could be compared to determine which was more responsive. Secondary analysis of other departmental datasets could be used to assess responsiveness of the same generic measures in other clinical contexts. The relationship between a scale's responsiveness and its construction, validity and reliability could be explored.

However, I felt this approach was too restrictive – being primarily focused upon the BRIDGE Study. It would not necessarily provide more general help to researchers wanting to know how to interpret the increasing amount of responsiveness data appearing alongside new and existing outcome measures. More fundamental questions therefore arose about what information is provided by responsiveness statistics; how these should be interpreted; and how do the statistics perform under varying conditions (for example, given a 'small' or 'large' sample size)? Similarly, what was the impact of different proposed ways of calculating these responsiveness statistics and how should they be incorporated into a research design?

Addressing these questions was more likely to produce an original and important contribution to the field. Therefore, I refined my focus to concentrate upon one of the most commonly reported responsiveness statistics (standardised response mean) and to assess its performance under different data distributions. How robust to varying distributional assumptions was this method - how prone to bias? The first component study of this thesis

addresses these questions by modelling this effect size statistic in a series of computer simulations (**presented in chapter 3**).

A second problem identified: An outcome measure through its basic design or application may be relatively insensitive to important change when it occurs – for example, if baseline scores exhibit ceiling effects when improvement is expected.⁷ Of course, there are many influences on our ability to detect clinically important change, for example, patient recall bias. Quality of life, one of the key outcomes of the BRIDGE and other studies, is primarily measured through respondent self-report. One's ability to comprehend, recall, judge and verbalise subjective states when answering a survey question may be affected by a variety of psychological mechanisms. This has increasingly been recognised in the growth of interest in cognitive approaches to survey methodology (CASM).⁸ Furthermore, and more fundamentally, it has been suggested that the very nature of subjective constructs such as quality of life may change over the course of time for an individual.⁹

Whether we have a responsive outcome measure or not, our ability to measure change is surely confounded if the very nature of the measured construct alters somehow during the course of a study independent of any intervention. How would we know if this is happening in a trial, is it possible to predict when it might occur, what would be the impact on our ability to detect 'true' change? What do observed changes in quality of life mean to those patients?

Assessing methodological options and opportunities: Whilst exploring the literature on responsiveness and developing ideas about the modelling work, I became aware of the emerging work of Sprangers and Schwartz who had been formalising a theory of response shift.¹⁰ These authors described a model by which an individual may amend their self-evaluation of their concept of quality of life as a result of a change in their personal internal standards of measurement, a change in their values or a re-definition of quality of life. Furthermore, they had summarised potential methods for assessing the occurrence of response shift.¹¹ I felt that the scope of my work should be

expanded to address these issues. At the same time I was commencing a new trial of the management of people with internal derangement of the knee (The DAMASK Study). Again, we aimed to develop an outcome measure for this study and there was potential to explore response shift in this study population. Reported clinical work on response shift had mainly taken place in oncological or palliative care settings. Would it be possible to identify response shift in patients with disabling but less serious conditions?

Of the various methods proposed for assessing response shift, I chose a primarily qualitative, interview-based approach for what would become the major component of this thesis. Recommendations for triangulation of method when studying response shift led me to develop a set of related component studies within this general theme of response shift. The use of an individualised method for measuring quality of life also provided the opportunity to explore the construct of quality of life for this study population. Furthermore, lengthy waits in accessing care for knee injuries fostered exploration of the impact of such delays on patient outcomes in the light of response shift. Together, the theme of response shift and quality of life studies are reported across a number of chapters following a common framework introduction (**chapters 5 to 12**)

Instrument and patient perspectives in assessing change: Both themes address the measurement of change. Responsiveness focuses upon the statistical representation of sensitivity to change as an attribute of an *outcome measure*. Response shift focuses upon the nature and process of changing response to subjective constructs as an attribute of the *individual patient*. Fundamentally, all effect size methods assume a common metric between baseline and follow-up scores. At the very least, re-calibration response shift would challenge this assumption, whilst re-conceptualisation could render a derived change score meaningless. Responsiveness assessment and response shift mechanisms both mediate our ability to detect and interpret change. Indeed response shift may fundamentally alter the nature of self-reported change prior to consideration of scale responsiveness. The response shift theme complements the responsiveness theme by exploring the validity

of assumptions and procedures used when assessing change in quality of life.

Perceived scientific and personal benefits: The chance to study response shift with this group of patients offered personal and scientific advantages. As an applied psychologist and health services researcher this provided a useful balance of opportunities in developing my research skills. Each component addressed change and its measurement from varying methodological perspectives. I hoped this would result in a broader and more rounded perspective to my enquiry. It would also afford me the opportunity to develop diverse research skills. I also believe strongly in the added value of combining different methodological approaches and the need for researchers to be receptive to (if not necessarily practitioners of) all suitable methods.

Thus the two interrelated themes of my thesis have developed from my experiences as a researcher and clinical trialists, and my encounter with unresolved methodological issues. They reflect a personal interest in research methodology by combining qualitative and quantitative approaches in what I believe is a fruitful synergy. I hope that they make an original contribution to current methodology in the field of outcome assessment, especially the evaluation of quality of life, and provide assistance to other researchers grappling with such issues in the future.

Contents	Page
Title page	i
Summary	ii
Declaration	iii
Acknowledgements	iv
Forward	vi
Contents	xii
Chapter 1 Study aims and thesis presentation	1
1.1 Theme one: Assessing the performance characteristics of the standardised response mean effect size	1
1.2 Theme two: Evaluating the nature and role of response shift in changes in self-reported quality of life in patients with suspected or confirmed internal derangement of the knee.	2
1.3 Synthesis	5
Chapter 2 Assessing change and scale responsiveness	7
2.1 Performance criteria for outcome measures	7
2.2 An increasing interest in responsiveness	9
2.3 Measuring the ability to detect change: conceptual and taxonomic approaches	12
2.4 Relating responsiveness, validity and reliability	22
2.5 The applications of responsiveness	26
2.6 Factors affecting a scale's sensitivity to change	29
2.7 Summary	32
Chapter 3 Assessing relative bias in responsiveness statistics: the standardised response mean	35
3.1 Introduction	35
3.2 Methods: development and pilot work	42
3.3 Study I: Assessing methods for handling negative change values, and effects of varying distributional parameters on SRM	50
3.4 Study II: Relative bias of SRM given a non-normal (skewed) distribution	61
3.5 Discussion	68
3.6 Summary	74

Chapter 4	Assessing change and response shift	77
4.1	The emergence and theoretical modelling of response shift	77
4.2	Mechanisms of response shift	83
4.3	A review of empirical response shift studies	87
4.4	Psychometric and clinical importance of response shift	96
4.5	Methods for assessing response shift	102
4.6	Summary and implications for thesis studies	107
Chapter 5	Research framework for evaluating response shift	109
5.1	Overview of study aims and presentation of response shift studies	109
5.2	Background to study	110
5.3	Response shift studies: design	115
5.4	Principal outcome measures	120
5.5	Assessing response shift using SEIQoL-DW	123
5.6	Interview schedules	125
5.7	Qualitative analytic approach	132
5.8	Summary of analytic plan	136
5.9	Summary	138
Chapter 6	Response shift interview study: overview and descriptive results	139
6.1	Summary of patients approached, consented and interviewed	139
6.2	Completion of baseline and follow-up interviews	142
6.3	Description of sample	145
6.4	Discussion	148
6.5	Summary	149
Chapter 7	A thematic analysis of quality of life content	150
7.1	Introduction	151
7.2	Methods	156
7.3	Results	158
7.4	Discussion	193
7.5	Summary	201
Chapter 8	Stability and change in quality of life: qualitative evidence for re-conceptualisation and re-prioritisation response shift from individualised assessment	203
8.1	Comparing cue profiles – a case study approach	203
8.2	A typology of change	228

8.3	Mechanisms contributing to response shift	232
8.4	Results synthesis and general discussion	249
8.5	Summary	253
Chapter 9	Evaluating scale re-calibration response shift using the retrospective pretest-posttest design	255
9.1	Introduction	256
9.2	Methods	260
9.3	Results	265
9.4	Discussion	283
9.5	Summary	290
Chapter 10	Exploring the relationship between cue level and cue weighting	291
10.1	Introduction	291
10.2	Methods	293
10.3	Results	294
10.4	Discussion	307
10.5	Summary	311
Chapter 11	Group level assessment of re-conceptualisation and re-prioritisation: a commentary and some recommendations	313
11.1	Introduction	313
11.2	Using aggregate SEIQoL data in quantitative analyses	314
11.3	Administration and cue elicitation	319
11.4	Using aggregate SEIQoL data to assess response shift	324
11.5	Recommendations for quantitative group comparisons	327
11.6	Summary	329
Chapter 12	Compounding the problem, confounding the outcome - waiting for knee care in South Wales	330
12.1	Introduction	330
12.2	Methods	334
12.3	Results	334
12.4	Discussion	344
12.5	Summary	348
Chapter 13	Synthesis and summary	349
13.1	Introduction	349
13.2	Responsiveness theme	349

13.3	Response shift theme	352
13.4	Responsiveness and response shift: making the connection	360
13.5	Conclusion and contribution	361
Appendices		364
I	Fortran (source file) program ia_na.f95 (Responsiveness modelling: Study I)	365
II	Fortran (source file) program ib_nav1.f95 (Responsiveness modelling: Study II)	378
III	Background publications from response shift literature review	392
IV	Patient approach letter (Radiology)	397
V	Patient approach letter (Orthopaedics)	398
VI	Patient information sheet	399
VII	Baseline interview schedule	402
VIII	Follow-up interview schedule	409
IX	Individual case summaries	421
X	Publications and presentations	461
Bibliography		462

Chapter 1: Study aims and thesis presentation

Abstract: *The thesis is presented according to its two constituent research themes: responsiveness and response shift. This chapter describes the aims and objectives of the studies that comprise the two themes. The studies are largely presented in autonomous chapters and these are indicated in this chapter. Finally, a synthesis of the thesis aims is presented.*

1.1 Theme one: Assessing the performance characteristics of the standardised response mean effect size

The ability of an outcome measure to detect clinically important change is referred to as responsiveness. It is not known how variation in underlying distributional parameters affects estimates of responsiveness. The *primary aim* of this research theme was to evaluate the performance of one of the most commonly reported effect size statistics – the standardised response mean – under differing distributional conditions. This was assessed by measuring the relative bias of derived statistical estimates. Specific objectives were:

- To model the impact of varying sample size, mean underlying effect size and shape of underlying distributions (normal and lognormal) to reflect conditions found in evaluative studies of quality of life
- To determine the importance and impact of different methods currently used (or proposed) for dealing with negative change values (i.e. those that move in the opposite direction of expected change)
- To compare the attributes of parametric effect sizes and their non-parametric alternatives

An introduction to responsiveness is presented in **chapter 2** and all empirical work under this theme is presented in **chapter 3**.

1.2 Theme two: Evaluating the nature and role of response shift in changes in self-reported quality of life in patients with suspected or confirmed internal derangement of the knee

The way that an individual appraises their own quality of life may change in response to ill-health, medical intervention or another catalyst. This change in self-evaluation is known as response shift. The *primary aim* of this research theme was to describe and assess response shift in the self-evaluation of quality of life in patients with suspected or confirmed internal derangement of the knee. Several specific objectives were identified:

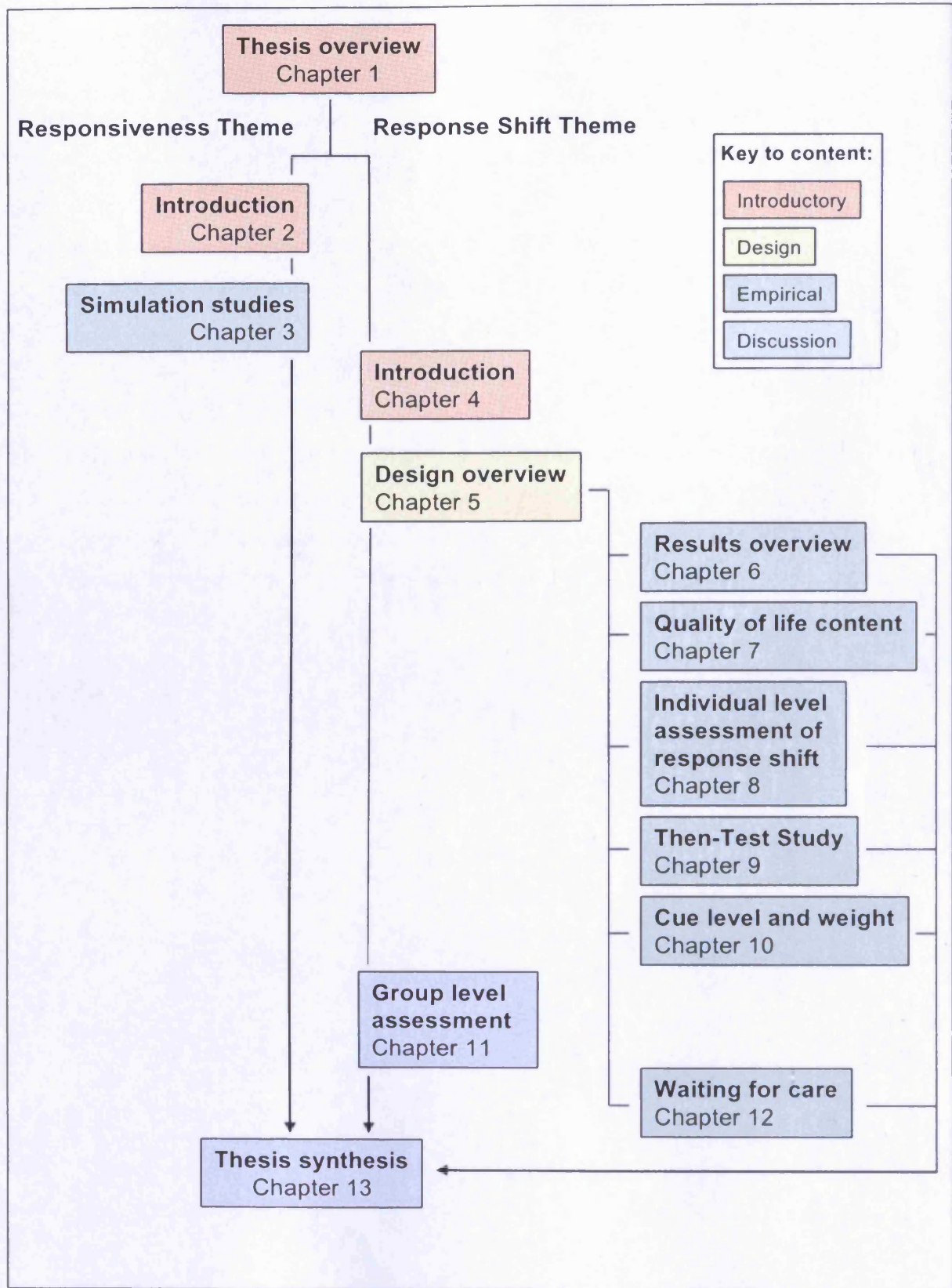
- To determine whether patients with a disabling but not life-threatening condition experience a response shift in their self-reported quality of life over a period of six months
- To explore which facets of response shift may have occurred (i.e. change in internal standards, values or re-conceptualisation)
- To explore individual characteristics that may affect response shift (including stage of clinical management, gender and age)
- To explore the mechanisms (behavioural, cognitive and affective processes) that may affect response shift in this patient population

A general introduction to response shift is presented in **chapter 4** and a description of the interview framework for the empirical studies of this theme is provided in **chapter 5**. The empirical work addressing the specific objectives above is presented in **chapters 6, 8, 9 and 10**. An additional three *secondary aims* were also specified:

- To explore and describe the nature of quality of life as perceived by patients with knee problems using an individualised approach to assessment (**Chapter 7**)
- To critically appraise methods for the quantitative analysis of response shift using individualised quality of life data (**Chapter 11**)
- To assess the impact of diagnostic and treatment delay upon quality of life for patients referred by their general practitioner with suspected internal derangement of the knee (**Chapter 12**)

A graphical overview of the structure of the presented thesis is shown in figure 1.2.1. The figure broadly indicates the nature of chapter content - introductory, design, empirical and discussion. Chapters shown as empirical generally integrate results with specific background literature, methods and discussion.

Figure 1.2.1 Map of thesis content



1.3 Synthesis

Modelling of the standardised response mean statistic and its variants will provide greater understanding of its merits and under what circumstances it might be used. Identification of systematic bias may indicate suitable correction factors and recommendations for when it may be optimally used.

Longitudinal studies that employ a relatively unresponsive outcome measure decrease the chance of detecting significant change and increase the chance of a Type II error. Larger trial samples are required to detect the same underlying effect – increasing response burden and research costs.

Potentially useful interventions may take longer to evaluate; to enter clinical practice; and thus, to benefit patients. Therefore, poorly chosen or performing measures present scientific, fiscal and ethical problems. Criteria for evaluating scale responsiveness need to be firmly established and their attributes better understood.

More fundamentally, the possibility of response shift may confound simple comparison of scores across two time points. What is being measured and its meaning for the individual may change over time. Assessment which draws upon that comparison may have to be adjusted with response shift in mind. This may affect not only assessment of responsiveness but also reliability. Furthermore, response shift raises questions about what is actually being measured and, therefore, the validity of the outcome measure itself.

Therefore, the thesis addresses two aspects of the measurement of change. The first, responsiveness, focuses upon the statistical representation of sensitivity to change as an attribute of an outcome measure. The second, response shift, focuses upon the nature and process of changing response to subjective constructs as an attribute of the individual patient. Responsiveness assessment and response shift mechanisms both mediate our ability to detect and interpret change. As response shift may fundamentally alter the nature of self-reported change (prior to subsequent consideration of scale

responsiveness) it is the main focus of the thesis. The response shift theme complements the responsiveness theme by exploring the validity of assumptions and procedures used when assessing change in quality of life. The thesis as a whole critically evaluates one method of assessing responsiveness - an area currently characterised by an absence of consensus; and assesses the nature and measurement of response shift in a longitudinal setting.

Preface to chapter 2

The following chapter provides a general introduction to the topic of responsiveness and reviews approaches to the assessment of change. The chapter sets the context for the empirical studies of the responsiveness theme presented in chapter 3. This work is finally summarised in the general thesis synthesis in chapter 13.

Chapter 2: Assessing change and scale responsiveness

Abstract: *The ability of a health outcome measure to detect change will determine its value for use in longitudinal studies. However, debate about what aspect of change to measure; the methods of its assessment and even the terms used to describe this attribute remain unresolved. This chapter introduces and reviews conceptual and taxonomic perspectives proposed for change measurement. A distinction between two broad approaches is reviewed. The first aims to correlate change scores with an external criterion or 'gold standard' of change. The second summarises the use of statistics to evaluate the extent to which outcome measures detect change. The performance characteristics of one such statistic - the standardised response mean (a form of effect size) will be formally evaluated through simulation modelling in chapter four. This chapter therefore provides a context for the empirical assessment of this commonly applied method for reporting a scale's responsiveness to change. Finally, the importance, application and determinants of scale sensitivity to change will be reviewed.*

2.1 Performance criteria for outcome measures

For researchers who are either embarking on the development of an outcome measure, or simply deciding between existing alternatives, there are a number of key questions that may be considered (box 2.1).¹² These questions address the performance characteristics of the outcome measure. Implicit within this framework are criteria of validity (ii), reliability (iii) and interpretability (iii). Terwee and colleagues considered validity to include the sensitivity of a measure to clinically important change (labelled as evidence of longitudinal validity).¹² The first question in this framework addresses the purpose to which the measure is being applied. Such purposes have been categorised by Kirshner and Guyatt into discriminative, predictive and evaluative.¹³

Box 2.1.1 Issues in developing or selecting an outcome instrument

- i Specification of measurement goals: what concepts are to be assessed?
- ii Does the instrument measure these concepts?
- iii How well does the instrument measure these concepts?
- iv How should the outcomes from the instrument be interpreted?

Adapted from Terwee and colleagues, 2003¹²

There are a number of established and emergent criteria for assessing the suitability and performance of health outcome measures.^{6 7 14-17} Frameworks with a particular focus upon quality of life assessment have also been presented. For example, Hays and colleagues proposed reliability and validity as the two psychometric considerations necessary for quality of life measures.¹⁶ Like Terwee and colleagues, the authors considered scale responsiveness of importance but essentially a component of validity assessment. In contrast, Testa and Simonson specified five properties: coverage, reliability, validity, responsiveness and sensitivity.¹⁴ Regardless of whether responsiveness is a component of validity or not, the importance of the concepts to which the term has been applied is becoming generally accepted.

In view of the various approaches described above, I have chosen to focus upon the classification of commonly considered criteria listed in the 1998 UK Health Technology Assessment review of patient-based outcome measures.⁷ The review addressed measures for use in clinical trials and the measurement of change due to interventions. Eight criteria by which such measures should be assessed were distinguished (table 2.1.1). More established psychometric criteria such as validity and reliability are described alongside less commonly described factors (e.g. appropriateness and interpretability), although no rank ordering is implied. Responsiveness is described as the sensitivity of an instrument to detect changes of importance to individuals. Fitzpatrick and colleagues note the heterogeneity of approaches to the assessment of responsiveness.

Table 2.1.1 Criteria for assessing use of patient-based outcome measures for use in clinical trials

Criteria	Issues addressed
Appropriateness	The match of an instrument to the specific purpose and questions of the trial
Reliability	Requirement that an instrument is reproducible and internally consistent
Validity	Judgement of whether an instrument measures what it purports to measure
Responsiveness	Whether an instrument is sensitive to changes of importance to individuals
Precision	The number and accuracy of distinctions made by an instrument
Interpretability	How meaningful are the scores from an instrument
Acceptability	How acceptable is an instrument for responders to complete
Feasibility	The extent of effort, burden and disruption to staff and clinical care staff arising from use of an instrument

Adapted from Fitzpatrick R and colleagues (1998)

2.2 An increasing interest in responsiveness

The relative novelty of responsiveness as a criterion considered of importance for outcome assessment is reflected in both the variety of methods used for its determination and also the definition of the concept. Terwee and colleagues identified and reviewed 26 definitions of responsiveness and 31 different measures.¹² They attributed this variation to the differing goals being addressed in each study reviewed. Beaton and colleagues identified sixteen different definitions of responsiveness, a number of which were provided by the same authors.¹⁸ Like Terwee, they also considered that the varying definitions reflected the different types of change being assessed. They constructed a taxonomy of responsiveness choosing an operational definition to encompass each type of change being described. Some approaches, for example, that of de Bruin and colleagues, view responsiveness as the ability to detect changes in the concept being measured.¹⁹ In comparison, the definition offered above by Fitzpatrick, also includes the concept of

importance of the change being observed.⁷ Other perspectives reflect more clinical aspects, for example, the detection of clinically important change.²⁰

Systematic literature search: To chart the increasing interest in responsiveness, I searched the Medline database for articles published from 1966 until the end of 2004. The search strategy involved identifying articles within two component sets (representing firstly, the criteria of responsiveness and secondly, health and quality of life outcome assessment). Abstracts of articles including both components were retrieved.

Thus, set 1 included articles with the keywords 'responsiveness to change' or 'sensitivity to change'. Set 2a included articles with the keywords or subject heading 'quality of life' or 'psychometrics'. Set 2b included articles with the subject headings 'outcome assessment (Health Care)' or 'outcome and process assessment' or the keyword 'outcome assessment'. Articles in either set 2a or 2b were combined (figure 2.2.1) The final sample of 321 abstracts was derived from English language publications formed by combining those identified by both sets 1 and 2 (i.e. 1 'and' 2).

Abstracts were reviewed to confirm relevance and whether the papers described broadly outcome validation studies (with or without responsiveness reported) or review / commentary papers. Prior to 1980 there were only two references identified, one describing a method based on variance ratios and the other the correlation of change scores. Only from the early 1990s have papers started increasing in incidence. The majority of recent papers selected are reporting scale development, comparison or validation exercises (table 2.2.1). Although the increasing incidence of papers reporting responsiveness may mainly reflect a rising interest in this concept, it is also possible that it also reflects a general increase in the development and use of subjective outcome measures.

Fig 2.2.1 Search strategy for responsiveness articles

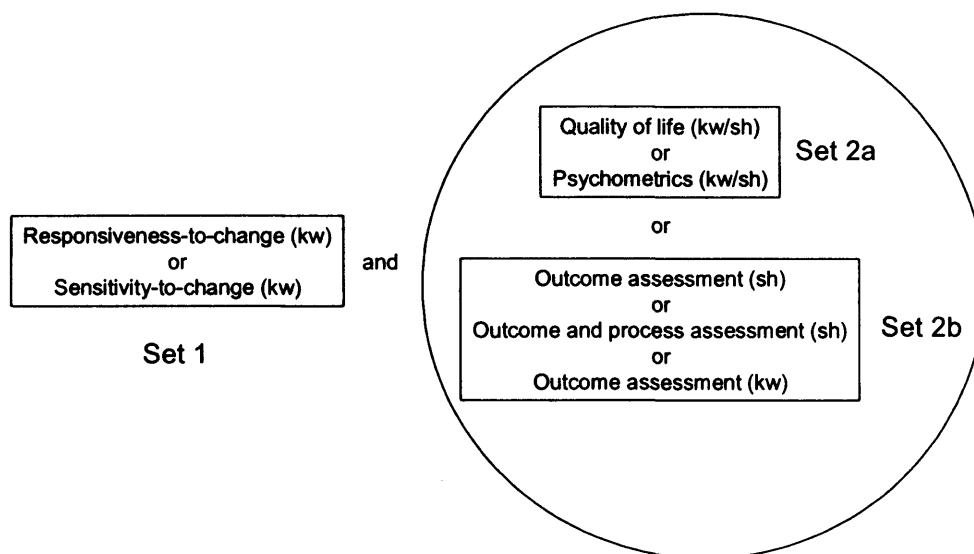


Table 2.2.1 Publications identified from Medline reporting or referencing sensitivity or responsiveness in quality of life outcomes

Year of publication	Nature of responsiveness reference			Total
	Validation	Validation / awaiting responsiveness	Review / commentary	
2004	40	1	6	47
2003	27	3	8	38
2002	24	3	6	33
2001	26	2	6	34
2000	15	2	7	24
1999	30	5	2	37
1998	19	1	5	25
1997	17	1	3	21
1996	10	1	2	13
1995	9	3	1	13
1994	4	-	-	4
1993	7	-	2	9
1992	4	-	2	6
1991	-	-	2	2
1990	1	-	-	1
1980-89	10	-	2	12
<1980	2	-	-	2
Total	245	22	54	321

2.3 Measuring the ability to detect change: conceptual and taxonomic approaches

Growing interest in responsiveness and a concurrent lack of consensus in concept definition has been reflected in the development of various frameworks addressing responsiveness (table 2.3.1). As noted above, some have sought to integrate responsiveness within a broader framework of criteria by which outcome measures may be judged. This has involved responsiveness being viewed either as a distinct psychometric criterion or located within the construct of validity. Other approaches have focused on the aims and methods of responsiveness assessment itself to delineate a distinct taxonomy.^{18 21} The following section highlights some the main approaches working from a historical perspective. Some authors have been included in table 2.3.1 even though their own focus may have been broader than simply responsiveness, for example, Stratford and colleagues.²² All authors have been included because they illustrate issues relevant to the ability to detect change.

Table 2.3.1 Classifications of relevance to responsiveness assessment and measures

Authors	Year	Focus	Structure / category	Definition	Comments
Deyo ²³	1986	Responsiveness as analogous to diagnostic test performance	Sensitivity to change Specificity	Ability to detect improvement Ability to discriminate between those who do/do not improve	Deyo's previous definition of sensitivity incorporated the concept of importance of the change ¹⁷
Lydick ²⁴	1993	Operational definitions of clinical meaningfulness	Distribution-based interpretation Anchor-based interpretation	Statistical distributions from a given study Change compared to other clinical changes	Most commonly represented by effect sizes <i>Anchors may be construct, discriminative or predictive references</i>
Liang ²⁵	1995	Responsiveness of clinical measures	Sensitivity to change Responsiveness	Statistical property of measure Judgement of the importance of change (heuristic)	
Fortin ²⁶	1995	Statistical approach to measuring change	Sensitivity Relevant change	The ability to detect change statistically (whether relevant or not) Change which is clinically meaningful	Sensitivity considered component of responsiveness

Table 2.3.1 (cont.) Classifications of relevance to responsiveness assessment and measures

Testa ¹⁴	1996	The assessment of quality of life outcomes	Responsiveness Sensitivity	Measure of association between observed change score and underlying construct Ability of measurement to reflect true changes or differences in underlying construct	
Stratford ²²	1996	Study designs for assessing meaningful change	Single group design Multiple group design		
Norman ²⁷	1997	Examine relationship between two forms of responsiveness effect size	Responsiveness-Treatment co-efficient Responsiveness-Retrospective co-efficient	Overall observed treatment effect Contrast changed and stable using external criterion (transition item)	Function of measure and treatment being evaluated
Murawski ²⁸	1998	Applications of responsiveness scores	Generalisability Interpretability Comparing measures Deriving sample size	Application of score to other settings Clarification of meaning of change score Choosing between available outcome measures	

Table 2.3.1 (cont.) Classifications of relevance to responsiveness assessment and measures

Husted ²¹	2000	Review and categorisation of forms of responsiveness assessment	Internal responsiveness	Ability to measure change over particular pre-specified time period	Function of measure and treatment being evaluated
			External responsiveness	Extent to which change in measure corresponds to change in reference	Property of measure only
Beaton ¹⁸	2001	Responsiveness as an attribute of scale specific to the context of a study	Who axis	Is information from the study to be analysed and interpreted at the individual or group level?	Two levels delineated: individual and group
			Which axis	Which scores are being contrasted?	Three levels delineated (between person differences at one time-point; within person change over time; between person differences of within-person change)
			What axis	What type of change is being quantified	Five levels delineated (minimum potentially detectable by instrument; minimum detectable given measurement error of instrument; observed change by instrument in population; observed change in population judged by external standard; observed important change in population)
Terwee ¹²	2003	Type of change able to detect	Change in general	All change regardless of relevance or meaning	Equated to 'sensitivity to change' (Liang ²⁵)
			Clinically important change Real change in underlying concept		Implies judgement of importance Requires judgement of importance and criterion measure of underlying concept

The detection of change and the importance of change

An early and enduring distinction has been made between the assessment of change as a statistical property of an instrument and the ability to identify change that is meaningful or important according to some external criterion.¹⁴

²³⁻²⁶ The inconsistent and overlapping use of terminology has contributed to the confusion and lack of consensus in this area. Thus the ability to detect important or meaningful change has been variously described as 'specificity', 'anchor-based interpretation', 'responsiveness' and simply 'relevant change'.²³⁻²⁶

In 1984, Deyo summarised performance criteria for functional outcome measures, describing sensitivity as the ability of a measure to detect change or differences of a magnitude that would be considered important.¹⁷

Subsequently, however, Deyo and Centor used the analogy of diagnostic test performance to describe specificity and sensitivity as different, but related, attributes of responsiveness (table 2.3.1).²³ *Sensitivity* was used to describe the ability of a measurement scale to detect improvement after an intervention of known efficacy. As this may give an incomplete picture of the usefulness of a scale, they described the other necessary attribute as *specificity* – the ability of the scale to discriminate between those improve and those who remain stable.

Clinicians' lack of familiarity with subjective outcome measures (compared to traditional biomedical measures) and an attempt to convey clinical meaningfulness led Lydick and Epstein to review methods for interpreting measurement changes.²⁴ Although their focus was on how such changes might impact upon an individual, they also summarised population-based perspectives. Operational definitions of clinical meaningfulness were classified as either distribution-based or anchor-based interpretations. The former involved the use of statistical distributions from a given study (e.g. the effect size statistic). The latter required the comparison of observed change with other measures of clinical change (e.g. patient global rating). The familiarity of such external anchors would make change in subjective outcomes more accessible to clinicians.

Fortin and colleagues defined responsiveness as the ability to detect clinically meaningful change - a responsive measure discriminates between clinically relevant and irrelevant change. Relevant change and sensitivity were thus distinguished with an external criterion required to determine the former and the latter being assessed statistically using effect sizes.²⁶ Responsiveness statistics were categorised as either sensitivity statistics or correlational methods assessing the ability to distinguish between relevant and irrelevant change. Liang made a similar distinction between the statistical property of a measure and a measure of the importance of a change. However, he chose to label these sensitivity to change and responsiveness respectively.²⁵ Liang considered that heuristic methods (patient or clinician reported transition items) should be used to determine the importance of the observed change.

Similar to Liang's approach, Testa and Simonson described a dichotomy between responsiveness and sensitivity.¹⁴ Responsiveness was described as the measure of association between change in the observed outcome measure score and the underlying construct being measured - its assessment requiring an external criterion. Sensitivity indicated the ability of the measure to determine true change in the underlying construct. For Testa and Simonson, sensitivity is critically determined by gradations of the measurement scale's metric.

Responsiveness and study design: Norman contrasted responsiveness coefficients based on overall treatment effects (standardised effect sizes and standardised response means) with those based upon retrospective categorisation of change (Guyatt's index of responsiveness).²⁷ The former are a function of both outcome measure and treatment. Norman asserts that the post hoc separation of respondents within the retrospective approach makes it conceptually different from the prospective method. Hence, although often reported together, the two approaches may lead to different conclusions. The theoretical and practical distinction between prospective and retrospective approaches is also addressed in chapter nine.

Taking a somewhat different perspective, Stratford and colleagues described a hierarchy of study designs (from single to multiple groups) and analytic approaches for evaluating change scores.²² Single group designs include simple before-after studies, and studies with a baseline run-in period to assess variance in 'stable state'. However, such designs are problematic because if no change is detected it may be unclear whether this is due to an absence of clinical change or rather the measure was unable to detect change. Furthermore, simple before-after designs do not allow for assessment of instrument performance on stable patients.

Therefore, Stratford and colleagues favour multiple group designs which they consider to represent a better estimate of change and stability allowing distinction between patients with varying levels of change. Such designs include placebo controlled randomised trials, quasi-experimental studies where non-randomised groups are followed up, and designs involving retrospective assignment to changed or stable groups on the basis of an external criterion of change. The ability to identifying a suitable external reference or 'gold standard' is a common criticism of the latter approach.

The use of an external criterion: Husted reviewed the properties and interpretation of commonly reported responsiveness statistics, distinguishing between *internal* and *external* approaches.²¹ Internal approaches address the ability to measure (largely group level) change over a pre-specified time period. External approaches relate change on an outcome measure to a criterion measure, (usually at an individual patient level). Most commonly reported measures of responsiveness fall under the internal category, including the widely used standardised response mean (an effect size statistic). However, as effect sizes are not externally validated they may not reflect important change at all. For this reason, Husted and colleagues considered a sole reliance on internal measures of responsiveness may result in an incomplete picture of a scale's usefulness. They favoured external approaches, particularly methods based on regression modelling.

A common challenge of external approaches is identifying a suitable criterion. Indeed, Husted and colleagues recognised that whilst scale change score and external criterion may be correlated, neither may be adequately related to the underlying construct of interest. The criterion measure also needs to demonstrate variability in response to enable evaluation based upon correlations. Finally, should the new outcome measure be novel and measuring a previously unevaluated construct, there is unlikely to be a suitable external standard available.

Assessing different types of change: Based upon the type of change being described in a study, Beaton and colleagues produced a formal taxonomy of responsiveness with three axes labelled *Who*, *Which* and *What*.¹⁸ The *Who* axis reflects whether results are being interpreted at either group or individual level. The second axis, refers to the timing of data collection (between and within person change). The third axis delineates what type of change is being assessed – broadly separating out for consideration magnitude of change and its importance. These specifically include minimum change potentially detectable; minimum change detectable given inherent measurement error; observed change in a given population; observed change in an ‘improved’ group; observed change in individuals with ‘important’ change. This taxonomy therefore, combines ideas about study design introduced by Stratford and inherent in other formulations. It also addresses the concept of change, important change and sources of external referents. The taxonomy emphasises that a scale’s responsiveness is an attribute that is specific to the context of the change being evaluated.

More recently, Terwee and colleagues categorised operational definitions of responsiveness, also based upon the kind of change a responsive instrument should be able to detect.¹² Their three categories are: (1) the detection of change in general regardless of its relevance or meaning; (2) the ability to detect clinically important change and (3) the ability to detect real changes in the concept being measured. The latter two categories both require a judgement about the importance of change. The third category, unlike the second, further specifies a gold standard. However, as the authors

themselves note, satisfactory gold standards do not exist for health-related quality of life, perhaps diluting the categorical distinction in practical terms.

A contrasting view is represented by Lydick and Epstein who regard quality of life responses as outcomes in themselves, and as such equivalent to disease states (rather than equivalent to laboratory measures of disease states). Thus they note the argument that any change in quality of life measure score is clinically important as this represents patient perception of change in health outcome.²⁴ However, patient perception of change in quality of life (for example, through self-reported transition items) and observed differences in reported levels of quality of life may not be equivalent.

How Terwee and colleagues assigned definitions of responsiveness to their three categories may be questioned. The 'real change' category includes the definition provided by Testa and Simonson.¹⁴ However, Terwee and colleagues have quoted the definition provided for sensitivity and not responsiveness. Testa and Simonson themselves do not provide an indication of what a gold standard for change would be, either for general or true change.

Within each of their three categories Terwee and colleagues note the gold standard being applied to determine general change, clinically important or real change. The source of this gold standard may be either due to treatment effect, or reported by either patient or clinician. The inclusion of a gold standard of real change in the third category is intended to differentiate it from the second category. It seems problematic, therefore, that the source used to define change (i.e. patient or clinician) shows considerable overlap between these two categories. Terwee and colleagues cite a number of studies where a gold standard of real change is apparently used. One such study using a patient-sourced gold standard of change is that of O'Keeffe and colleagues.²⁹ A patient-reported transition item is used to categorise patients into changed and stable groups. In this study the transition item specifically addressed cardiac-related health. The extent to which such an item represents a gold

standard for cardiac-related quality of life is debatable. Whilst one would expect the two to be correlated, they are not equivalent constructs.

The practical difficulty of inferring 'real change' is further demonstrated in a study cited by Terwee under their third category of responsiveness assessment. Deyo and Centor aimed to assess responsiveness in measures of functional status.²³ Responsiveness is reported as the correlation between outcome measure change scores and clinical variables. The clinical variables include patient rating of pain improvement, professional rating of improvement as well as other measures such as spine flexion. It is not clear that any of these clinical measures could be taken as gold standards for the construct being measured. Indeed, the paper's own authors note that no gold standard for functional status exists.

Beaton, Hogg-Johnson and Bombardier used a global transition item which they describe as a criterion for 'clinically estimated improvement' in health.³⁰ This is more akin to the second category of responsiveness rather than the third under which it was classified by Terwee. The transition item is used to classify patients prior to assessment of responsiveness using standardised effect size and standardised response mean – considered by Terwee and colleagues to be representative of the second category of responsiveness assessment.

Summary of issues: The formulations described above have addressed a number of related concepts in the measurement of change – a dichotomy between the simple assessment of change and the determination of the meaning of the observed change; the role of different study designs for the assessment of change; the use and choice of external criteria; the usefulness and application of different forms of responsiveness assessment and classification of types of change being described. There is an increasing sophistication as one proceeds through each new specification but perhaps the key concept is best captured by Husted in separating out internal and external approaches.

2.4 Relating responsiveness, validity and reliability

Whilst sensitivity to change is widely accepted as an important measurement characteristic of an evaluative outcome tool, there is continuing debate as to whether it is simply a component of validity.³¹ Whilst the criteria presented in table 2.1.1 consider validity and responsiveness separately, Fitzpatrick and colleagues note that other authors have integrated the two constructs.^{7 22 32 33} For example, De Bruin and colleagues comment that 'the ability to accurately assess changes ... in a longitudinal setting can be viewed as part of the validity of an evaluative instrument'.¹⁹

Kirshner and Guyatt distinguished between responsiveness and validity, and emphasised the need to establish responsiveness for evaluative instruments.^{13 34} Furthermore, Guyatt and colleagues state that a scale could be responsive but lack validity if, for example, it measures content irrelevant to the intended purpose of the instrument.³⁵ They illustrate this with reference to a responsive measure (Eastern Co-operative Oncology Group Criteria) intended to measure health status but which includes laboratory-based measures. Hays and Hadorn refuted this suggestion by pointing out that the example simply demonstrates that a measure may perform well on one test of validity but not on another.³¹ They also note that a measure valid at one time point should be valid at another.

As described above, Husted and colleagues separated out internal measures of scale sensitivity to change and external responsiveness (relating scale change scores to an external criterion of change).²¹ They argue that the latter is distinct from longitudinal construct validity. In their view, studies of longitudinal construct validity select measures to establish both convergent and discriminant validity. In establishing external responsiveness, the external measure is selected to represent an accepted indicator of change. However, this assessment can still be regarded as one component of longitudinal construct validity (i.e. convergent validity).³¹ Indeed, identical approaches may

be used for determining responsiveness in longitudinal studies and validity in cross-sectional studies.^{12 36}

That differing goals of responsiveness assessment have led to conflicting views of the relationship between 'responsiveness', validity and reliability may be illustrated by a comparison between two conceptualisations provided by Terwee and colleagues and Guyatt and colleagues respectively.^{12 37} Guyatt and colleagues described two essential scale properties: validity and a high signal to noise ratio (table 2.4.1). The latter is assessed using either reliability or responsiveness co-efficients depending upon whether the assessment is cross-sectional or longitudinal. What constitutes signal or noise again depends upon timing and purpose of assessment. Within this framework reproducibility and responsiveness are not independent criteria - they are manifestations of the same dimension in two different contexts.

Table 2.4.1 Relating responsiveness to reliability and validity: a comparison of two approaches

Author	Dimension	Criterion / method	
		Cross-sectional	Longitudinal
Terwee ¹²	Validity		Responsiveness: Ability to measure change in construct <i>Method:</i> I: Correlation with external criterion of change II: Treatment effect sizes (within pre-specified hypotheses)
	Reliability		Longitudinal reproducibility <i>Method:</i> Treatment effect sizes (SRM)
Guyatt ³⁷	Validity		
	Signal to noise ratio	Reliability Signal: inter-subject variability Noise: intra-subject variability	Responsiveness Signal: intra-subject variability Noise: inter-subject variability

Terwee also locates the measurement of treatment effect (specifically the standardised response mean) under longitudinal assessment of reliability – where it is labelled longitudinal reproducibility. However, the purpose of assessing a scale's ability to measure real change is referred to as longitudinal validity. This is assessed by relating change scores of the scale and an external criterion of change (correlational approach). Guyatt suggests the index of responsiveness as an example of responsiveness (signal-to-noise ratio). Although this index attempts to relate change in part to an external criterion, their focus is upon determination of treatment effect (experimental approach). By clarifying these two scientific approaches (correlational and experimental) they have demonstrated how 'responsiveness' may address either validity or reliability. However, they do note that there is still a limited role for effect sizes to play in longitudinal construct validity if used with adequate prior specification of expected change.

The correlational approach to responsiveness, such as advocated by (for example) Husted and colleagues emphasises the location of responsiveness within the realm of measure validity.²¹ It seems reasonable to consider the ability to detect clinically important change as part of the overall validity of an instrument that purports to measure such changes. In this respect, the perspective of Terwee and colleagues seems to have credit, not least because it explicitly accommodates the use of both treatment effect sizes and approaches based on correlating change scores with external criteria.¹² Whilst these authors felt that effect sizes were only of limited value, they nevertheless retained such approaches as a means of determining longitudinal validity. Although the problem of classification may remain disputed, it does not critically impact on the perceived value of the construct. The term 'responsiveness' continues to be increasingly used and distinguished from other psychometric criteria. Perhaps though, it may become generally regarded as simply another form of outcome measure validation, albeit with a distinct label and methods.

An holistic approach to validation would entail the ongoing accumulation of multiple sources of information which together clarify whether a scale measures what is intended to measure.^{31 38} Assessing the validity of a scale designed to measure change would encompass ascertainment of responsiveness as part of that process. This would not preclude the gathering of other evidence of a scale's validity, for example, content validity.³¹ As a scale's validity is relative to its context and proposed application it is entirely consistent for conclusions about cross-sectional and longitudinal validity to be different.¹²

2.5 *The applications of responsiveness*

The calculation and reporting of sensitivity to change statistics may serve a number of purposes. In advocating standardised effect sizes, Kazis and colleagues described their role in providing a benchmark for interpreting change in trials; for comparing traditional biomedical with newer health status measures; comparing two drugs in separate trials and providing a more complete picture of clinical meaning in changing health status.³⁹ Interestingly, not included amongst these aims was performance comparison of different subjective scales. However, they commented that effect sizes may facilitate understanding of the relative sensitivity of traditional biomedical and health status measures.

Using efficient outcome measures that are more responsive to clinically important change will increase the chance of identifying therapies that improve patient well-being and reduce the number of patients required to demonstrate an underlying change.^{26 40} The ratio of sample sizes necessary to detect a clinical effect is inversely proportional to the square of the ratio of standardised response means:⁴¹

$$\frac{n_1}{n_2} = \left(\frac{SRM_2}{SRM_1} \right)^2$$

Liang illustrates this relationship by stating that a three-fold increase in scale responsiveness equates to a nine-fold reduction in sample size.⁴¹ Either developing new measures, or choosing between existing measures that are more sensitive to change will reduce patient burden in clinical trials and other longitudinal studies and reduce overall research costs.⁴² Therefore, the derivation and reporting of responsiveness statistics is key when selecting between available outcome measures.

Husted and colleagues argue that effect sizes are limited by their specificity to each study from which they are derived and that approaches which attempt to correlate changes score with an external criterion are more appropriate when selecting measures.²¹ Rather than for generalisation across clinical settings, the use of effect sizes have been recommended only for extending from pilot to full studies.²⁸ Even this restricted application would help reduce respondent burden, minimise multiple comparisons and guide sample size determination. Murawski and Miederhoff found variation in the effect sizes for outcome measures across applications, leading them to conclude that responsiveness was not an inherent characteristic of a scale. Whilst other factors such as size of true treatment effect and disease category have been shown to be influential, Wiebe and colleagues challenge this conclusion.⁴³ They found specific measures consistently superior to generic when used concurrently in studies with a true treatment effect, across heterogeneous disease groups.

Interpretability: Effect size statistics have been advocated as a means for understanding change on health status measures, especially for clinicians who may be less familiar with such instruments compared to biomedical outcomes.^{24 39} Demonstrating clinical significance may also be important for such subjective outcomes which may be viewed as softer and less clinically meaningful.²⁴ Benchmarks for interpreting the importance of effect sizes proposed by Cohen are now commonly cited in studies reporting scale sensitivity.³⁹ However, Kazis and colleagues comment that the appropriate classification of effect sizes may differ from those provided by Cohen. Whilst the same benchmarks have been used for interpreting different effect size

statistics (e.g. standardised effect size and standardised response mean) the appropriateness of doing so has also been questioned.^{21 33 40}

Cohen describes the effect size, a unit-less statistic, as an index of departure from the null hypothesis.⁴⁴ Operational definitions are provided for the qualitative descriptions of *small*, *medium* and *large* about which researchers are concerned when contemplating sample size calculation. That the basis for the benchmarks as reasonable but arbitrary and no more reliable than his 'own intuition' was acknowledged by Cohen.^{44 45} When two populations are being compared the null hypothesis can be stated as the difference in the value of the relevant parameters being zero. Cohen's benchmarks are shown in table 2.5.1. For example, a *large* effect size is intended to represent differences that are grossly observable – the example provided by Cohen was the difference in height between 13- and 18-year-old girls.

Table 2.5.1 Effect size benchmarks provided by Cohen for comparing means of two populations

Effect size	Description	% non-overlap of two population distributions	Proportion of combined variance accounted for by group membership
0.2	Small	14.7%	1%
0.5	Medium	33%	5.9%
0.8	Large	47.4%	13.8%

In summary, outcome measures which are more sensitive to change have many advantages from a scientific, ethical and economic perspective. Whilst the use of effect sizes to make generalisations about a scale's performance across clinical settings is debatable, they can still guide study planning and development. Although benchmarks for interpreting the clinical significance of effect sizes are commonly applied, their common application to different statistics (standardised effect size and standardised response mean) has been questioned.²¹ Finally, it should also be remembered that Cohen provided the benchmarks as a rule of thumb and other study factors should be considered with their use.

2.6 Factors affecting a scale's sensitivity to change

Instrument sensitivity may be compromised if the signal of underlying change is too weak or if noise (sources of measurement error) is too large.^{25 36 46}

Large subject variability in baseline scores may also reduce measurement sensitivity. Factors likely to play a role in instrument sensitivity include scope (i.e. whether general or specific); item and scale construction; stability of the target construct; mode of application and response style effects.

Generic and specific measures: Several studies have compared the sensitivity to change of generic and specific outcomes measures.⁴⁷⁻⁵⁰ For example, Garratt and colleagues found specific measures to be more sensitive than generic in patients with low back pain.⁴⁷ Murawski and Miederhoff reviewed 324 studies which had employed two or more measures concurrently and concluded that (disease-)specific measures were consistently more responsive.²⁸ Whilst this analysis was criticised for amongst other things the inclusion of non-randomised and uncontrolled studies, a more recent review of 43 randomised controlled trials reached the same conclusion.⁴³ Greater sensitivity of specific measures is influenced by their focus upon aspects of health of relevance to respondents and that they directly address the intervention being evaluated.

Choice of response scale: Item construction involves consideration of both response scale and item stem. Whether the underlying model for assessment is categorical (i.e. seeking to identify differences between cases and non-cases) or dimensional (i.e. assuming an underlying continuum) will affect sensitivity to change.⁶ The former focuses upon the presence or absence of attributes (e.g. symptoms) and may be represented by binary scales. The latter represents data on continuous scales. Outcome measured using binary variables provides less potential for change than with visual analogue or likert-like scale.⁵¹ Guyatt and colleagues compared the responsiveness of the latter two types of scales in a study of patients with lung disease and found no

significant difference.⁵² Interestingly the mean change was greater when using the VAS but so also was the variance. Whilst the optimal number of steps on the response scale to ensure responsiveness is not clear, the number of steps chosen will be also influenced by other considerations such as ease of completion and reliability.^{6 53}

Ceiling and floor effects: If a measure shows ceiling or floor effects, underlying change may not be evident in the change score.^{14 46} For example, Rouf and colleagues compared the responsiveness of three outcome measures in 112 patients with ankylosing spondylitis.⁵⁴ The measure with a greater spread of baseline scores, and fewer subjects at the high end of the distribution, was more responsive.

Stem construction: Vermeersch and colleagues highlighted question stem construction as a mediator of sensitivity, for example, if the time-frame covered by an item extends back further than the period of an intervention.⁵⁵ This would have the effect of diluting treatment effect by introducing more noise into the assessment. Items asking specifically about change may also limit sensitivity. For example, items in the Chronic Pain Grade confounds current state with patient assessed change - the last two scale items rating change in disability over the previous six months.⁵⁶ Paradoxically, patients with maximum disability would show better health on the CPG when re-assessed six months later compared to patients whose ability has deteriorated over a similar time period but whose absolute level of disability was not as severe.⁵⁷

Patient preferences: Incorporating patient preferences in scale construction has been suggested to promote sensitivity to change. Bessette and colleagues found that weighting items according to patient preference resulted in a more sensitive scale compared to an un-weighted version in patients with carpal tunnel syndrome.⁴⁸ However, individualised approaches that allow patients to select and weight items (e.g. SEIQoL) have not always been found to be more responsive compared to standardised measures.^{58 59} Other factors such as adaptation to disease and response shift may mitigate the

responsiveness of outcome measures thus reducing the advantages of individual weighting.⁶⁰

Number of items: The number of items forming a summary scale score may affect sensitivity. Moran and colleagues gradually reduced the number of items comprising each sub-scale of the Chronic Respiratory Questionnaire and found adequate scale responsiveness with only two items per scale.⁶¹ However, other scale properties were effected and reducing items increased the sample size required by about 10% in several sub-scales. Katz and colleagues compared short and long form generic health status measures in patients undergoing hip arthroplasty. They concluded that shorter forms were as responsive as longer measures.²⁰ However broader generalisable conclusions would benefit from a greater accumulation of evidence.

Choice of items: Item selection and reduction is another area where scale sensitivity may be affected. Van der Heijden and colleagues used clinicians' and researchers' judgement to select from candidate items.⁶² Potential items were rated in terms of how sensitive to change they were likely to be. Similarly, Montgomery and Åsberg developed a depression rating scale designed to be sensitive to change.⁶³ From an item bank, those that showed the greatest change before and after different courses of established treatment were selected.

Scale items for assessing static constructs may not be useful for measuring change.³⁸ The construct being measured may be relatively stable, for example, personality traits.⁵⁵ The level of true underlying change over time may be small making it difficult to detect change. If a scale contains a heterogeneous item pool assessing relatively stable as well as dynamic aspects, then its sensitivity to change may be affected. However, it is the stability of the target construct itself rather than the sensitivity of the outcome measure that is the limiting factor here.

Other influences: Factors other than scale design have also been investigated. For example, Chambers and colleagues assessed whether

different modes of data collection (self-completion, telephone or face-to-face interviews) resulted in different levels of responsiveness.⁶⁴ They found that mode of administration did not effect the size of change score detected using a global measure of health status. Response style effects may also play a role if they reduce the range of available response, within that provided by a scale. Montgomery and Åsberg describe how this may occur due to response styles such as central tendency bias.⁶³

2.7 Summary

The developing interest in change assessment is evident in the increasing reference to scale sensitivity or responsiveness to change in the medical literature. The continuing lack of either a commonly accepted definition of responsiveness, or a method of measuring it, may reflect differing types of change being addressed by researchers. Perhaps for the same reasons there is an according lack of consensus about how responsiveness relates to other performance criteria such as validity and reliability. Nevertheless, the importance of a scale that can detect change (whether clinically important or real) is not in dispute and there is gathering evidence of what may increase a scale's performance in this regard.

Amongst the conceptual approaches to change assessment presented, two issues are of particular relevance within this thesis. Firstly, clarity about the type of change being assessed is vital. The hierarchy of study designs presented by Stratford and colleagues helps to emphasise the distinction that should be drawn when comparing measures of responsiveness from different studies.²² Subsequent formulations by Beaton and colleagues and then by Terwee and colleagues again seek to clarify further the type of change being evaluated.^{12 18} Not doing so could lead to inappropriate conclusions when comparing responsiveness statistics measured both within and between studies.

Secondly, the binary clarification of internal and external measures of responsiveness exemplified by Husted and colleagues clearly identifies the role of an external criterion of change.²¹ Their framework usefully summarises similar ideas from previous formulations. Although the authors favour the external (correlational) approach, they have also summarised the limitations and scope of both internal and external methods. The availability and suitability of an external criterion remains a challenge for responsiveness assessment. This may be even more so for evaluation-based constructs such as quality of life.⁶⁵

The performance and associated relative bias of one responsiveness statistic will be evaluated in the next chapter. However, factors potentially affecting scale sensitivity were reviewed and may play an important role even before statistical evaluation commences. Fundamentally, the construct being assessed must itself be amenable to change. Furthermore, responsiveness is more likely if the measured construct lies on a theoretical continuum rather than a dichotomy. Of the relevant scale specific factors, the focus of the instrument (for example, generic or specific) and whether there are ceiling or floor effects appear to have a strong influence upon responsiveness. Interesting avenues for future investigation may include the role of data collection mode effects and how they interact with various response biases. In summary, features of the measured construct, scale design and study design may all have important influences upon responsiveness before consideration is given to choice of responsiveness statistic.

The first research theme in this thesis evaluates the standardised response mean, one example of an effect size and representative of the experimental approach to change assessment. Although the precise role of treatment effect sizes has been questioned, even those who favour the alternative correlational approach, recognise its use in assessing longitudinal reproducibility, interpretability and even longitudinal construct validity. The statistic is commonly used. It is also variably used – for example with either the whole sample or only those reporting improvement being used in its

calculation. There has been little consideration of how the statistic performs under these and other, distributional circumstances.

Preface to chapter 3

The current chapter provides the general introduction to the empirical work presented in the following chapter. The work models the performance characteristics of the standardised response mean (SRM) under different conditions to provide some certainty within this component of a diverse and contentious field.

Chapter 3: Assessing relative bias in responsiveness statistics: the standardised response mean

Abstract: Outcome measure responsiveness is increasingly being quantified using effect size statistics. One of the most commonly reported is the standardised response mean (SRM). The validity of this statistic in terms of the relative bias of its estimate has yet to be determined. Bias was explored in a series of simulation models within which sampling parameters and computational method were varied. The results are presented for initial piloting and model development, Study I (varying computational method) and Study II (varying underlying distribution). In Study I, either reversing the sign of negatively valued cases, or omitting them altogether, led to a substantially different SRM than when all values were included in the calculation. When all values were used and data were modelled from an underlying normal distribution, the amount of bias was generally small, even when varying the size and standard deviation of the underlying true change score. Relative bias was larger in smaller samples but was still no greater than 8%. In Study II a non-parametric version of the SRM showed little bias when used with data drawn from a log-normal distribution. However, using the more commonly reported parametric SRM resulted in substantial bias, especially when the sample size and underlying SRM were small. Overall the studies show that different approaches to handling negative values result in markedly different statistics, making comparison of SRMs across different studies and study designs potentially misleading. Furthermore, the less commonly used non-parametric SRM should be applied when reporting skewed data to avoid inflation of reported SRM and therefore, the risk of a Type I error.

3.1 Introduction

Effect size statistics: Chapter two described the increased interest in assessing responsiveness, and a broad distinction between external methods which compare change scores to a criterion of change and internal methods based on treatment effect. Internal methods include the paired t-test and

various effect size statistics, for example, the standardised effect size (SES), standardised response mean (SRM) and index of responsiveness (IoR) and formulae for their calculation are described in box 3.1.1.²¹ Each statistic directly addresses the size of observed change score (numerator) relative to a measure of variation (denominator). Effect sizes therefore, reflect a standardised ratio of signal to noise.³⁰ Different effect sizes have often been compared within the same study, partly reflecting uncertainty about which is the most appropriate index to use.^{49 50 54 66-68} This confusion is reflected in the different terminology used and occasional errors in calculation. For example, Feise and Menke described their use of the 'standardised response mean' but provide a description (and reference for) the standardised effect size.⁵¹

Box 3.1.1 Formulae for effect size statistics

$$SES = \frac{\bar{D}_x}{SD(X_1)}$$

$$SRM = \frac{\bar{D}_x}{SD(D_x)}$$

$$IoR = \frac{\bar{D}_x \text{ changed}}{SD(D_x) \text{ stable}}$$

\bar{D}_x = mean change score

X_1 = baseline score

$SD(D_x)$ = standard deviation for change score

Changed = improved / deteriorated sub-group

Stable = stable sub-group

The standardised effect size: Kazis described the calculation and use of an effect size for measuring health status change.³⁹ In this thesis, the statistic will be referred to as the standardised effect size, but is also known simply as an effect size.^{21 39} The denominator for this statistic is the standard deviation of the baseline score. The rationale for its calculation is that baseline scores are used as a proxy for a control group. An outcome measure with large baseline variability compared to mean change score will thus have a small effect size.²¹

As Kazis and colleagues were interested in size of change rather than statistical significance, they used the standard deviation of the baseline scores rather than the standard deviation of the difference between means. This has led to criticism that it does not incorporate response variance in its calculation.²⁰

The standardised response mean: A matched pairs effect size was introduced by Cohen for use in evaluative studies.^{69 70} The denominator used is the standard deviation of the change scores.^{25 40} An outcome measure with large variation in change scores relative to mean change score will thus have a small effect size.²¹ The statistic was subsequently renamed the standardised response mean (SRM) by Liang and has become widely used.^{21 36 40} As for the standardised effect size, high scores indicate a greater sensitivity to detect change.²⁶ As the SES does not contain information about the accuracy of an outcome measure in detecting change, the SRM has been viewed as superior.¹⁹ An early example of its application was provided by Anderson and Chernoff who used it for comparing outcome measures in rheumatoid arthritis.⁷¹ Like other effect size statistics, it has been inconsistently labelled, for example, being referred to as a Responsiveness-Treatment coefficient and an efficiency index.^{21 27 71}

The SRM is related to the paired t-test but uses standard deviation rather than the standard error of the mean as the denominator. It is thus, less influenced by sample size.^{30 40} Husted and colleagues described the relationship:²¹

$$SRM = \frac{t_0}{\sqrt{n}}$$

Variations in how SRM may be calculated and a number of distributional factors that may influence estimated SRM are now considered further.

Calculating effect sizes: Terwee and colleagues described the use of the SRM to detect change in general and to detect clinically important change.¹² The former may include all individuals in a sample (following receipt of

treatment of known efficacy), whilst the latter may include only those who have changed according to an external criterion of change. Therefore, other means of identifying 'changed' individuals have been sought.

Three studies demonstrate these two different approaches. In a method comparison study, Beaton and colleagues calculated effect sizes (SES and SRM) for patients with musculoskeletal disorders who were expected to show improvement between health status assessments (those in acute stage of natural healing and those in sub-acute stage in receipt of physiotherapy).³⁰ However, their principal analysis used an additional criterion of self-reported improvement to select those for inclusion in the calculation of effect sizes. When SRMs were re-calculated using the whole sample, regardless of response to the transition item, effect sizes were between 45-60% the magnitude of those based upon the self-reported improvement. Similarly, in a study of patients being treated for endometriosis-associated pain, Jones and colleagues calculated SESs for patients before and after surgery.⁷² Effect sizes calculated for the whole sample were substantially smaller than those calculated for those only reporting improvement on a transition item. Thus, effect sizes of 0.9 and 1.8 were found for whole sample and improved patients respectively on a pain sub-scale (as a benchmark, 0.5 is considered a medium effect size, 0.8 is considered large).⁴⁴ Finally, Schmitt and Di Fabio described the use of a prospective global disability rating scale to identify improved patients for whom effect size statistics were calculated.⁷³

Norman and colleagues described how the mean change for improved and deteriorated groups (categorised on the basis of transition item) may be aggregated if they are approximately equal.²⁷ Aggregation includes reversal of the sign for the deteriorated group. This is then used for calculating the IoR effect size (which the authors referred to as retrospective-responsiveness). The authors noted how the combination of improved and deteriorated cases is performed only after the means and standard deviations of the two groups have been affirmed.

The same data handling has also been reported for SRMs. For example, Brazier and colleagues calculated SRMs for two groups of patients with osteoarthritis of the knee - those attending rheumatology clinic and those about to undergo total knee replacement (TKR).⁷⁴ The SF-36 health transition item was used to categorise rheumatology patients into those who had *improved*, *remained stable* or *deteriorated*. Only those reporting change were used in the calculation of SRM and for those reporting deterioration their scores were reversed. It is not clear however, whether this was done only for those with a negatively signed change score or for all patients reporting deterioration regardless of actual change score. A single SRM is therefore provided for this rheumatology clinic sample. In calculating the SRM for patients in the TKR group (who were expected to improve), all cases were included, again with no details provided about the direction of observed change.

Expected effect size: As discussed in chapter two, Cohen described benchmarks for effect sizes based upon independent samples (treatment and control groups).⁶⁹ Notwithstanding the appropriateness of their application to SRM, the benchmarks have been widely applied and, for example, indicate an effect size of 0.5 as medium. Wiebe and colleagues reviewed the reported responsiveness of generic and specific quality of life instruments used in 43 randomised controlled trials.⁴³ They found a mean 'weighted' effect size of 0.57 for specific instruments and 0.39 for generic instruments in studies with a non-zero therapeutic effect. In their review, effect sizes were calculated based on study group assignment, hence differing from simple before and after calculation of standardised effect size.

Murawski and Miederhoff also compared generic and specific measures and calculated SESs for treatment groups in 39 studies.²⁸ They found a mean effect size of 0.66 for disease specific measures and 0.47 for generic measures. Marx and colleagues described the range of SRMs found in orthopaedic research using validated instruments - between 0.9 and 1.9.⁷⁵ In specific studies the size of effect may be much greater. Thus, Berber-Westin and colleagues reported effect sizes (SES and SRM) for sub-scales of the

Cincinnati Knee Scale in patients following ACL reconstruction ranging from 0.69 to 3.49.⁷⁶

Sample size: Marx and colleagues described the lack of sample size specification in studies of responsiveness, stating how authors have tended to use the sample size from reliability studies.⁷⁵ Whilst effect sizes have often been reported on large data sets, there are numerous examples of only small samples being used in their calculation. This is more likely to occur when sub-groups are identified for analysis on the basis of transition items. The number of cases available for effect size calculation may also be reduced by missing data, a particular problem when calculating change scores.⁷⁷ Examples of reported small samples include Ruperto and colleagues who included 26 children in study of active treatment for juvenile arthritis.⁶⁸ The authors further categorised this sample into 'improved' and 'not improved' according to a priori criteria. Fitzpatrick and colleagues calculated SES using sample sizes of 30 improved and 37 deteriorated patients according to a transition item.⁷⁸ Finally, Walker used samples of eight and twelve patients in calculating SES.⁷⁹

Shape of change distribution: In proposing the use of the SES, Kazis and colleagues noted the importance of baseline distribution characteristics.³⁹ They recommended using median and inter-quartile range to calculate effect sizes when outcome measure scores are highly skewed. This may well be the case for both patient and population samples for health status measures such as SF-36, where certain sub-scales may be inherently non-normal.⁸⁰ Whilst some studies report effect sizes calculated using median and inter-quartile range, Fitzpatrick's review noted that the use of non-parametric methods has seldom been employed.^{7 68 81}

Although a baseline distribution may be skewed, this may not be the case for the distribution of change scores. The shape of the change score distribution may not be easy to determine from study reports which often report only summary statistics (mean and standard deviation) for change scores. However, if an underlying normal distribution of change scores is sub-divided

(for the purpose of calculating effect size) according to study group or transition item, the resultant distribution may well not be normal.

Study aims: Therefore, to determine the impact of computational method, and to assess how the SRM performs under different distributional conditions, two studies were planned. A series of simulations were designed to model the impact of varying parameters of sample size, mean underlying effect size and shape of underlying distributions (normal and non-normal) upon the bias of effect size estimates. The main outcome therefore, was the bias in estimated SRM relative to the underlying (or 'true') SRM. These factors were explored across the two studies, each of which assessed variations in calculating SRM:

Study I: *Aimed to determine the impact of methods for calculating SRM which differ in how they handle negative change values (i.e. simulating deteriorated cases). Aimed to compare the relative bias of parametric and non-parametric SRMs given a normal underlying distribution.*

The method of calculating SRM reflects the type of change the statistic is aiming to represent. Including all values as generated (both positively and negatively signed) reflects assessment of change in general. Selecting sub-groups on the basis of an external criterion reflects assessment of clinically important change. The true value of treatment effect therefore is changed in the process. In a study with an overall positive effect, the true SRM for an 'improved' sub-group will inevitably be larger than for the whole sample, assuming a good correlation between criterion and outcome measure. A wholly valid external criterion (e.g. transition item) would be perfectly correlated with change score and the regression line would intercept both axes at zero. All negative change scores would correspond with self-ratings of deterioration and all positive change scores would correspond with improvement.

The modelling approach in study I assumes such a correlation between change score and external criterion. Modelling the three methods of

calculating SRM (using all values as generated, reversing the sign for negative change scores and omitting negative change scores) is expected to result in different SRMs. The study is therefore expected to demonstrate and emphasise the nature of the difference.

Study II: Aimed to compare the relative bias of parametric and non-parametric SRMs given a non-normal (skewed) underlying distribution

3.2 Methods: development and pilot work

Modelling of the distributions associated with the SRM statistic was explored using computer simulation. Using simulation ensures that parameters of the underlying distribution are known, thus allowing assessment of the relative bias in SRM estimation compared to a known 'true' value. Simulation programs had to be developed to model the parameters of interest and tested for accuracy and validity. This section describes the process of piloting and initial development work.

Pilot study: Method

Pilot simulations were written in Fortran 90 under a UNIX operating system (on a mainframe computer). The initial values randomly generated by the modelling programs represented change scores. The SRM was calculated by the program on the basis of these generated values. Each simulation program produced 1000 samples and for each sample calculated the corresponding SRM statistic. Output data files from each simulation were opened in SPSS and the mean value for SRM calculated. Pilot simulations used change scores drawn from a normal distribution. The effect of varying the sample size, mean change score and standard deviation upon the value of the SRM was modelled in three series of simulations (table 3.2.1). Relative bias was calculated for each simulation – that is the degree to which the observed SRM generated deviated from that specified within the modelling program. Relative bias was defined and calculated as follows:

$$R. \text{ bias} = \left(\frac{\text{estSRM}}{\text{trueSRM}} \right) * 100 - 100$$

Table 3.2.1 Overview of pilot simulations

	Modelling parameter		
	Change score	sd	Sample size
Series 1	10	2	15 to 1000
Series 2a	0.5 to 10	2	15
Series 2b	0.5 to 10	2	100
Series 3	10	1 to 50	100

In the first series (comprising seven simulations), the mean change score and standard deviation were held constant at 10 and 2 respectively (i.e. a ‘true’ SRM of 5) for each simulation. The size of the sample was varied using the following values: 15, 25, 50, 100, 250, 500 and 1000.

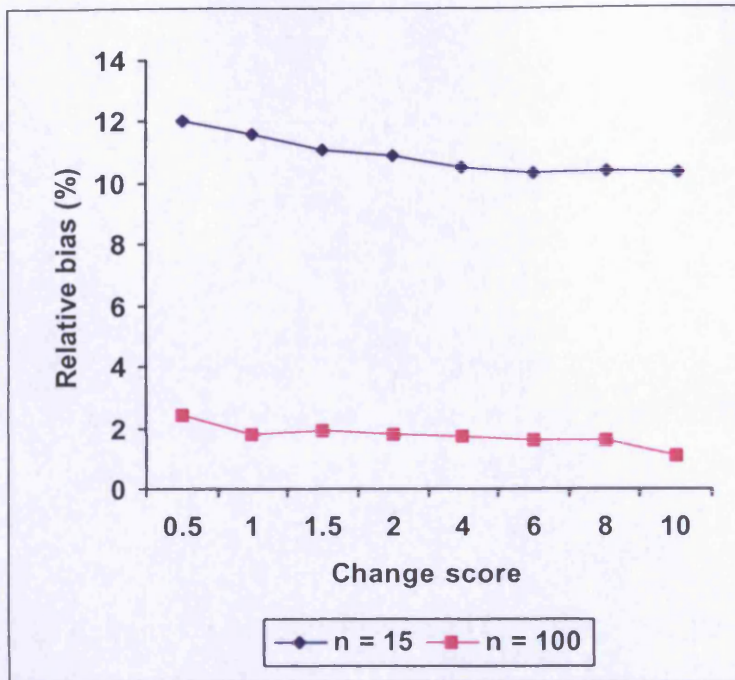
In the second series of simulations, the sample size and standard deviation were held constant whilst the mean change score was varied. Two sets of simulations were run, both with a standard deviation of 2. In the first set sample size was held at 15, whilst in the second set, the sample was increased to 100. In both sets of eight simulations the mean change values used were 0.5, 1, 1.5, 2, 4, 6, 8 and 10 (i.e. to simulate SRMs ranging in value from 0.25 to 5).

In the third set of simulations, the mean change score and sample size were kept constant (10 and 100 respectively) whilst the standard deviation was varied using values 1, 2, 4 and 50 across four simulations.

Pilot study: Results

In series 1, a sample size of 15 (mean change 10, standard deviation 2) resulted in a relative bias of 8.6%, whilst in a large sample (n=500) the bias was 0.2%. In series 2, varying the size of change score and holding sample size and standard deviation constant resulted in less variation in relative bias with 12% to 10.3% (sample size of 15) and 2.4% to 1.1% (sample size of 100). A plot of the relative bias from series 2a and 2b is shown in figure 3.2.1 and emphasises the impact of sample size on relative bias.

Figure 3.2.1 SRM: Relative bias plots for small and large samples with varying change score



In the third series of simulations the largest bias detected (5%) was in the model with a large standard deviation (50). In all simulation models, detected bias was positive and usually small.

Developing the main modelling programs

Following initial piloting, for practical reasons model development was switched to PC-based FORTAN software (Salford FTN95).⁸² This allowed greater ease in program editing compared to the line editor available on the mainframe; faster and easier program compilation and submission; and superior file management capability. Furthermore programs could be run from the local PC without submission to remote mainframe. However, the NAG routines previously available on the mainframe were no longer accessible using the PC-based software. Such routines were used to generate normally distributed pseudo-random numbers in non-repeatable sequences for use in all simulation models. Therefore, a program to replace these NAG routines had to be developed.

Generating random numbers: In pilot modelling, the syntax for creating a random distribution from the uniform distribution used the FORTRAN intrinsic function RANDOM_NUMBER. This intrinsic sub-routine produces numbers in a uniform distribution over the range:

$$0 \leq x < 1$$

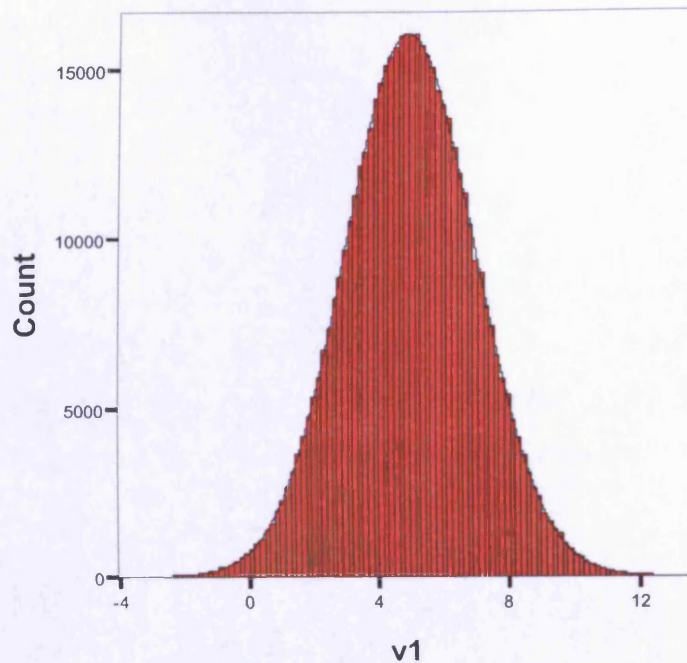
The original syntax also used a separate sub-routine to re-set the seed to obtain a non-repeatable sequence. The new program used the Salford FTN95 (run-time library) routine RANDOM to return a pseudo-random double precision value. This routine produces a uniformly distributed double precision (D) random number x such that:

$$0.0D0 < x \leq 1.0D0$$

As this routine would produce the same sequence of numbers each time it is run, the SALFORD FTN95 intrinsic routine DATE_TIME_SEED@ was used to select a new 'seed' for the number generator. The seed is set to a value based on the current DATE/TIME (drawing upon the computer CPU) to obtain a non-repeatable sequence.

A new test program (seed1.exe) was written which produced 500,000 normally distributed pseudo-random numbers with mean 5 and standard deviation 2, from the uniform random number generator. The program was run three times and output directed to separate text files which were then read into SPSS. Histograms for each distribution were produced to verify the shape of the distributions (figure 3.2.2). Every 50,000th value was listed (n=10 values) to verify the uniqueness of the three distributions (i.e. to confirm that the program was not producing the same sequence of values for each sample it generated).

Figure 3.2.2 Verifying random number generation: histogram of output values ($n=500,000$; mean = 5; sd = 2) from one run of FORTAN model seed1.exe



Number of valid simulation samples: One method for handling negatively signed values (representing 'deteriorated' cases) when calculating effect size omits them from the calculation. Calculating a non-parametric effect size requires at least four valid (hence positive) values to estimate the median, and upper and lower quartiles. The developing simulation program counted and omitted the number of simulations where this wasn't the case.

Understandably, this was more likely to occur in simulations with a smaller modelled sample size and with a smaller modelled effect size. Invalid simulation samples were found only when modelling SRMs 0.1, 0.25 and 0.5 and only with sample sizes of 15. Out of 10,000 simulations for these three SRMs there were respectively only 68, 19 and 1 simulation samples with less than four valid values. Therefore, the modelling program was adjusted to take account of a varying number of valid simulation samples.

Simulation size: The pilot modelling programs had been run using 1,000 simulations. The pattern of distributions produced showed some

inconsistencies, especially when modelling smaller SRMs. Therefore, test program ia_nd (using an SRM of 0.8) was run five times to compare the shape of the distribution of relative bias values. Using a sample size of fifteen, the smallest relative bias value resulting was 3.44 and the largest 6.11 (table 3.2.2; figure 3.2.3). Therefore, the number of simulations was increased to 10,000 and the program again run five times (table 3.2.3). This time there was much less variation in relative bias score produced and a much more consistent pattern of relative bias distributions (figure 3.2.4). All programs thereafter were run using 10,000 simulations.

Table 3.2.2 Relative bias of SRM using 1,000 simulations (SRM = 0.8)

	Sample size used in simulation					
	15	30	45	90	180	360
Run 1	5.7859	3.3874	1.3192	0.7875	0.1920	0.0277
Run 2	3.4442	2.4832	2.3804	1.1921	0.4756	0.1406
Run 3	5.7578	2.5104	1.5148	0.7224	0.6679	-0.0618
Run 4	5.5838	2.7756	0.9700	1.0630	0.1148	0.2565
Run 5	6.1154	2.6507	0.7057	0.3530	-0.1305	0.2659

Figure 3.2.3 Relative bias of SRM using 1,000 simulations (SRM 0.8)

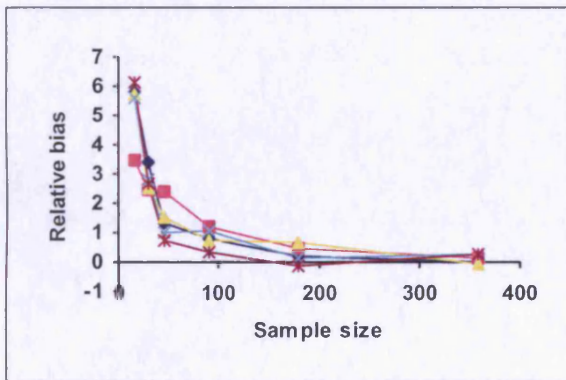
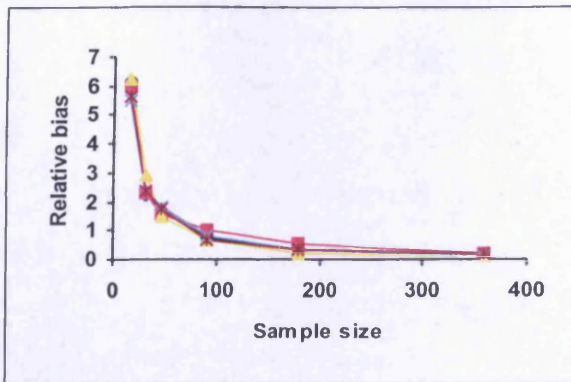


Table 3.2.3 Relative bias of SRM using 10,000 simulations (SRM = 0.8)

	Sample size used in simulation					
	15	30	45	90	180	360
Run 1	6.1686	2.2624	1.7014	0.7540	0.3478	0.2040
Run 2	5.9514	2.3436	1.5934	1.0101	0.5208	0.2285
Run 3	6.2358	2.9364	1.4823	0.7004	0.2273	0.1330
Run 4	5.5011	2.4052	1.8407	0.8447	0.3291	0.1415
Run 5	5.6507	2.3574	1.7969	0.7022	0.3107	0.1725

Figure 3.2.4 Relative bias of SRM using 10,000 simulations (SRM 0.8)

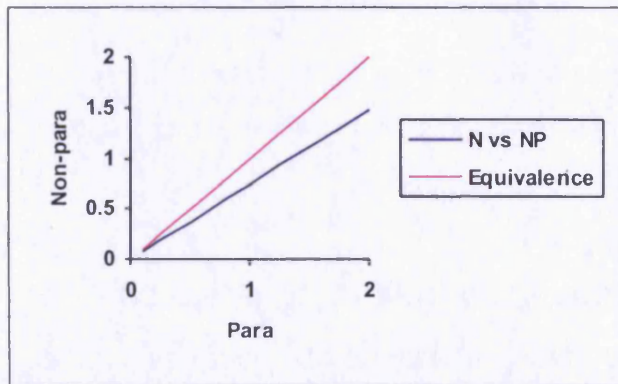


Parametric and non-parametric SRMs: Whilst the 'true' parametric SRM is specified in the modelling program (by setting the mean change and standard deviation), this is not the case for the 'true' non-parametric SRM. A non-parametric SRM may be calculated by dividing the median change score by the inter-quartile range of the change distribution. The parametric SRMs to be modelled in the current studies are shown in table 3.2.4 alongside their derived non-parametric equivalents, and plotted in figure 3.2.5. Calculating a non-parametric SRM in this way results in a smaller reported SRM – approximately three-quarters the size of the parametric SRM. So, for example, a 'true' parametric SRM of 1.0 equates to a 'true' non-parametric statistic of 0.74 for a normal distribution.

Table 3.2.4: Modelled SRMs: parametric and non-parametric equivalent

Parametric	Non-parametric
0.1	0.07413009
0.25	0.18532522
0.5	0.37065043
0.8	0.59304069
1.0	0.74130086
1.25	0.92662606
1.6	1.18608138
2.0	1.48260172

Figure 3.2.5 Parametric and non-parametric equivalent SRMs (including line of equivalence)



Model parameters: The following values were used for the parameters within the two modelling studies:

Sample size: Pilot modelling used individual sample sizes in the range 15 to 1,000. A small sample was associated with a much greater degree of relative bias. In practice, studies reporting responsiveness seldom include very large samples. Therefore, for the main modelling study the following steps were used:

N = 15, 30, 45, 90, 180, 360

Effect size: In a simulation study of the relationship between distribution- and anchor-based approaches to interpreting changes in health-related quality of life, Norman and colleagues used effect sizes from 0 to 1.5 (with increments of 0.25).⁸³ In the current studies the change score was kept constant at 1 throughout all modelling programs whilst the standard deviation was varied to produce SRMs from 0.1 to 2.0 (table 3.2.5). Specific standard deviations were chosen to produce SRM to no more than two decimal places.

Table 3.2.5 Modelled standard deviation and SRM

	A	B	C	D	E	F	G	H
SRM	0.1	0.25	0.5	0.8	1.0	1.25	1.6	2.0
sd	10	4	2	1.25	1	0.8	0.625	0.5

Sampling distribution: Values drawn from two underlying distributions will be used: normal and log-normal.

Specific methods and results are presented below for each study in turn. Discussion for both studies is presented in a combined section at the end of the chapter.

3.3 Study I: Assessing methods for handling negative change values, and effects of varying distributional parameters on SRM

Methods

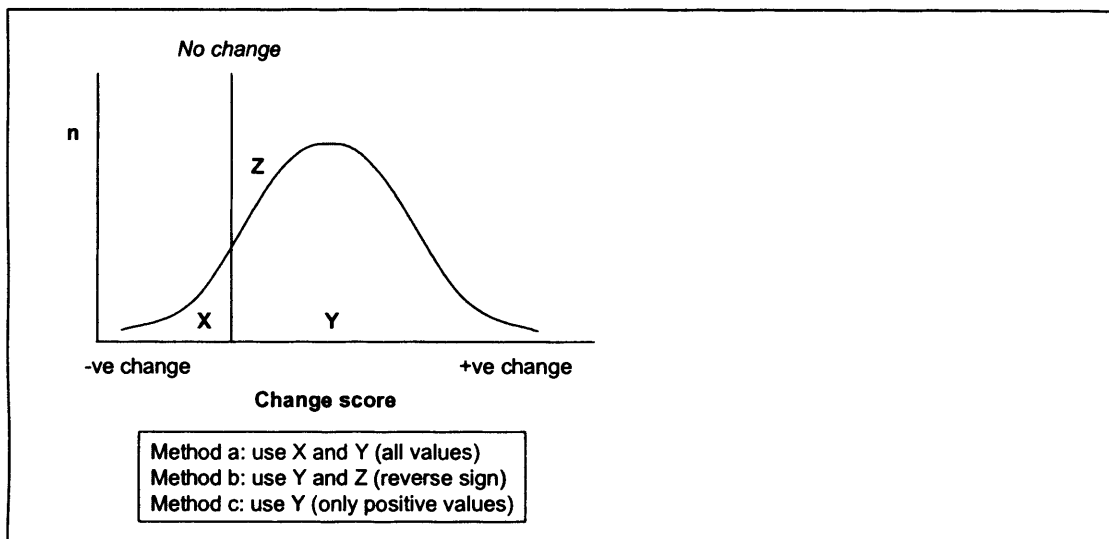
The first aim of study I was to assess the impact of different methods for calculating SRM (table 3.3.1), specifically dealing with negative values, upon the estimated SRM. This was modelled under optimum conditions (i.e. with values drawn from a normal distribution). The second aim was to compare relative bias of parametric and non-parametric SRMs given a normal underlying distribution. A single underlying distribution of change scores was assumed. The three computational methods for SRM compared were:

Table 3.3.1 Methods for calculating SRM in simulation study I

	Method	Purpose
a)	<i>Including all values as generated</i>	Simulates the calculation of SRM using all cases to detect change in general (following treatment of known or assumed efficacy)
b)	<i>Including all values but with the sign reversed for originally negative values</i>	Simulates the aggregation of improved and deteriorated sub-groups (for example, identified by self-reported transition item) where scores for deteriorated sample are reversed
c)	<i>Including only positive values</i>	Simulates the calculation of SRM using only cases where improvement has been observed (for example, though the use of self-reported transition item)

The three approaches are compared graphically in figure 3.3.1. Using all values as generated includes sections X and Y of the distribution. Reversing the sign of negative values increases the frequency of observations in the lower end of the resulting positive distribution – effectively contributing to section Z. Using positive values only includes just section Y of the distribution.

Figure 3.3.1 Methods of handling negative values when calculating standardised response mean



Relative bias will be computed within the modelling program (for method a only) as a measure of the difference between estimated and 'true' SRM. The calculations in methods b and c inevitably change the nature of the 'true' SRM – each method will result in a different 'true' SRM. The optimal method of calculating SRM from the three modelled will be chosen for further modelling in study II. Both parametric standardised response means (pSRM) and non-parametric standardised response means (npSRM) were calculated. The modelling plan is summarised in table 3.3.2:

Table 3.3.2 Study I modelling plan

		Sample size (a-f)					
		15	30	45	90	180	360
SRM (A-H)	0.1	N(Aa)	N(Ab)	N(Ac)	N(Ad)	N(Ae)	N(Af)
	0.25	N(Ba)	N(Bb)	N(Bc)	N(Bd)	N(Be)	N(Bf)
	0.5	N(Ca)	N(Cb)	N(Cc)	N(Cd)	N(Ce)	N(Cf)
	0.8	N(Da)	N(Db)	N(Dc)	N(Dd)	N(De)	N(Df)
	1.0	N(Ea)	N(Eb)	N(Ec)	N(Ed)	N(Ee)	N(Ef)
	1.25	N(Fa)	N(Fb)	N(Fc)	N(Fd)	N(Fe)	N(Ff)
	1.6	N(Ga)	N(Gb)	N(Gc)	N(Gd)	N(Ge)	N(Gf)
	2.0	N(Ha)	N(Hb)	N(Hc)	N(Hd)	N(He)	N(Hf)

Note: N denotes normal distribution being modelled.

Each program cell will model the three calculation methods.

Each program includes all sample sizes for one SRM (i.e. per row) – therefore there are eight programs in total.

Results

pSRMs: The estimated *pSRM* when using all values as derived (method a) was very similar to the modelled SRM (table 3.3.3). However, the values of the estimated *pSRMs* using method b (reversing the sign of negative values) and method c (dropping negative values) were very different. This was especially the case with SRMs under 1. For example, a modelled SRM of 0.8 and sample size of 90 resulted in a *pSRM* of 1.40 and 1.56 using methods b and c respectively – (a corresponding inflation of 175% and 195%).

Table 3.3.3: Estimated pSRMs using methods a, b and c for each modelled SRM and sample sizes

Sample size	Method	Modelled SRM							
		0.1	0.25	0.5	0.8	1.0	1.25	1.6	2.0
15	a	0.11	0.27	0.53	0.84	1.05	1.33	1.70	2.12
	b	1.39	1.39	1.40	1.46	1.52	1.64	1.87	2.19
	c	1.51	1.51	1.55	1.64	1.71	1.82	2.02	2.29
30	a	0.11	0.26	0.51	0.82	1.02	1.28	1.64	2.06
	b	1.36	1.36	1.37	1.42	1.49	1.60	1.82	2.14
	c	1.41	1.44	1.49	1.59	1.66	1.77	1.97	2.23
45	a	0.10	0.25	0.51	0.81	1.02	1.27	1.63	2.03
	b	1.35	1.34	1.36	1.41	1.48	1.59	1.81	2.12
	c	1.39	1.41	1.48	1.57	1.65	1.76	1.95	2.22
90	a	0.10	0.25	0.50	0.81	1.01	1.26	1.61	2.02
	b	1.33	1.34	1.35	1.40	1.47	1.58	1.80	2.11
	c	1.36	1.40	1.46	1.56	1.64	1.75	1.94	2.20
180	a	0.10	0.25	0.50	0.80	1.01	1.26	1.61	2.01
	b	1.33	1.33	1.34	1.40	1.47	1.58	1.79	2.10
	c	1.35	1.39	1.45	1.55	1.63	1.74	1.93	2.19
360	a	0.10	0.25	0.50	0.80	1.00	1.25	1.60	2.00
	b	1.33	1.33	1.34	1.40	1.46	1.58	1.79	2.09
	c	1.35	1.38	1.45	1.55	1.63	1.74	1.92	2.19

The differences between the pSRMs produced using the three methods are demonstrated graphically in figure 3.3.2 (for a sample size of 90). Only at larger SRMs are differences between method a, and methods b and c reduced. Even using the largest sample size there remains a small difference between the three methods. Methods b and c produce similar SRMs across the range modelled, although those produced by method c are always slightly larger. The pSRMs derived using methods b and c also are relatively insensitive to increasing modelled SRM and cover a much narrower range than for method a. The same pattern of pSRMs is found using smaller and larger sample sizes (figures 3.3.3 to 3.3.5).

Figure 3.3.2 Observed pSRMs using each calculation method, by modelled SRM: sample size of 90

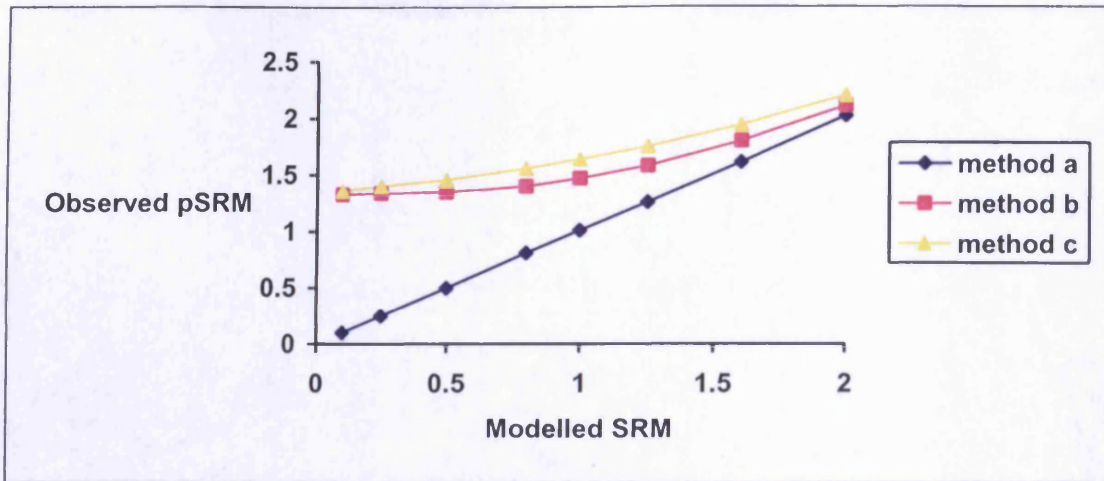


Figure 3.3.3 Observed pSRMs using method a for each sample size

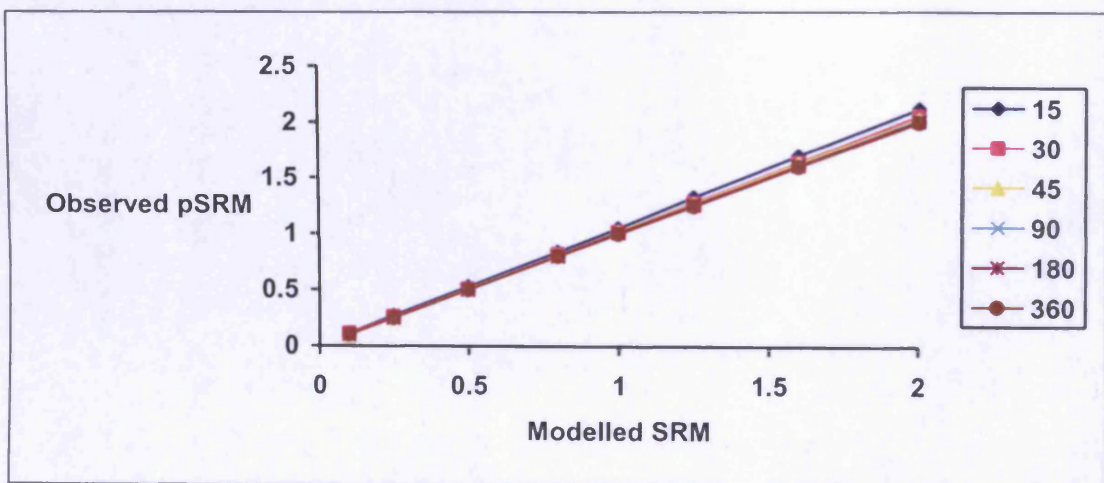


Figure 3.3.4 Observed pSRMs using method b for each sample size

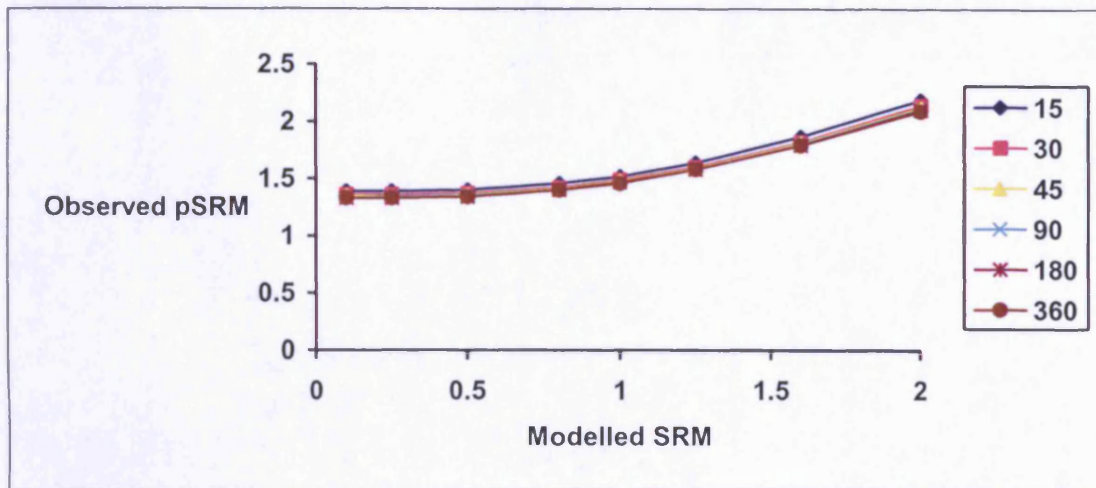
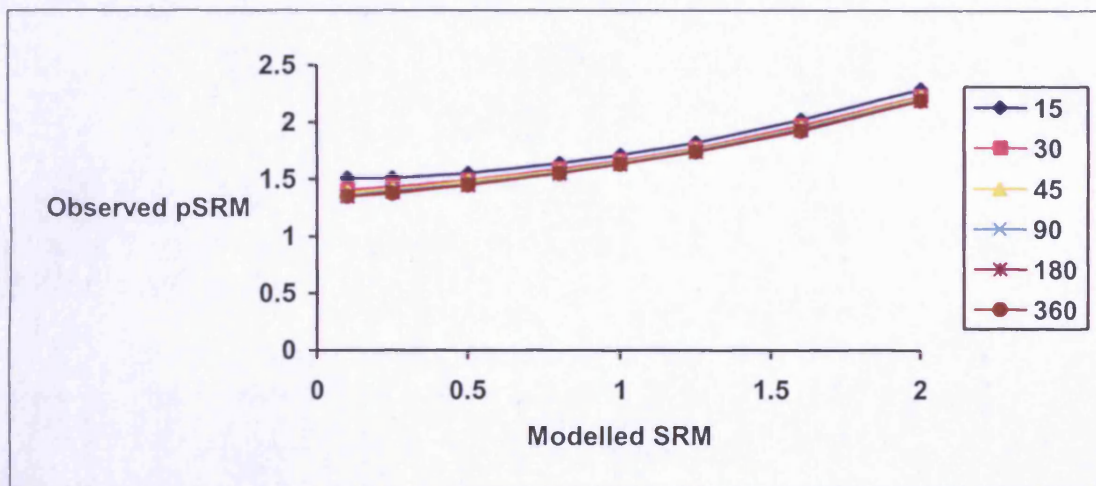


Figure 3.3.5 Observed pSRMs using method c for each sample size

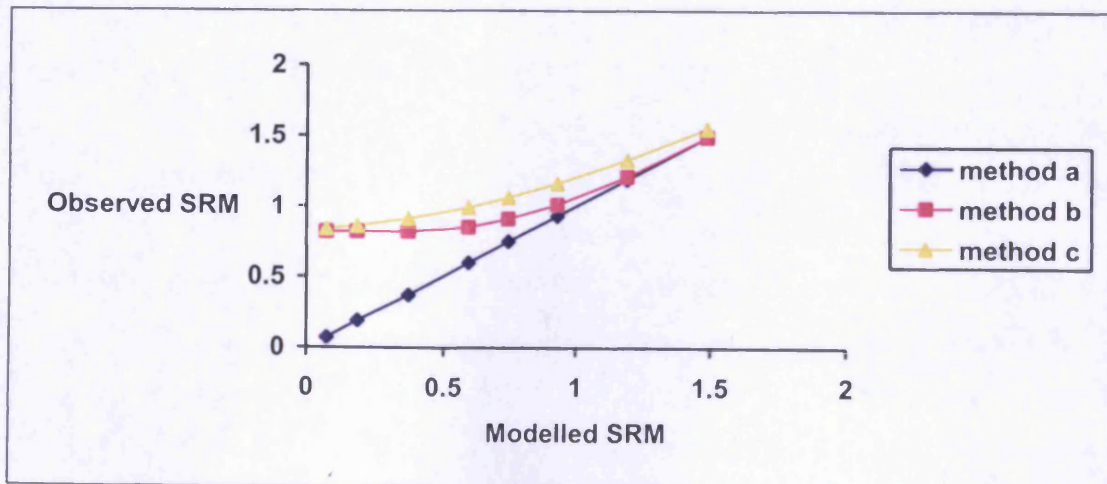


npSRMs: The *npSRM* was derived using each calculation method (still using values drawn from an underlying normal distribution). The pattern of *npSRM* distributions by sample size and modelled SRM was similar to that found for their parametric equivalents (table 3.3.4). Estimated *npSRMs* derived using method a were similar to the 'true' SRM. Those produced by methods b and c were substantially larger – demonstrated in figure 3.3.6 for a sample size of 90. This was particularly so with 'true' SRMs of 1 or less. Estimated *npSRMs* using method b (sign reversed) were comparable to those produced by method c (negative values dropped). Nevertheless, *npSRMs* produced by method b tended to be slightly smaller in the mid-range of modelled SRMs compared to method c (figure 3.3.6).

Table 3.3.4: Estimated npSRMs using methods a, b and c for each modelled SRM and sample sizes

Sample size	Method	'True' npSRM							
		0.074	0.185	0.371	0.593	0.741	0.927	1.186	1.483
15	a	0.08	0.19	0.36	0.61	0.77	0.96	1.23	1.53
	b	0.86	0.86	0.86	0.89	0.94	1.05	1.25	1.53
	c	0.88	0.90	0.95	1.03	1.09	1.19	1.36	1.59
30	a	0.08	0.19	0.38	0.60	0.75	0.94	1.20	1.50
	b	0.83	0.83	0.83	0.87	0.92	1.02	1.21	1.50
	c	0.86	0.88	0.93	1.01	1.07	1.17	1.33	1.56
45	a	0.08	0.19	0.37	0.60	0.75	0.93	1.19	1.49
	b	0.82	0.82	0.83	0.86	0.91	1.01	1.21	1.50
	c	0.85	0.87	0.93	1.00	1.07	1.16	1.33	1.56
90	a	0.07	0.19	0.37	0.60	0.75	0.93	1.19	1.49
	b	0.82	0.82	0.82	0.85	0.91	1.01	1.21	1.49
	c	0.84	0.86	0.91	0.99	1.06	1.16	1.33	1.55
180	a	0.07	0.19	0.37	0.60	0.74	0.93	1.19	1.49
	b	0.81	0.81	0.82	0.85	0.90	1.00	1.21	1.49
	c	0.83	0.86	0.91	0.99	1.06	1.15	1.32	1.55
360	a	0.07	0.19	0.37	0.59	0.74	0.93	1.19	1.49
	b	0.81	0.81	0.82	0.85	0.90	1.00	1.20	1.49
	c	0.83	0.86	0.91	0.99	1.05	1.15	1.32	1.55

Figure 3.3.6 Observed npSRMs using each calculation method, by modelled SRM: sample size of 90



To assess the relative impact of the constituent components upon the estimated SRM, the mean change and standard deviations are shown in table 3.3.5 (for a sample size of 90). The mean change scores observed using methods b and c are inflated (i.e. originally set at 1) and are much greater than for method a at the lower SRM levels. For higher SRMs the observed values for methods b and c approach the original true modelled change score. The standard deviations for method a are underestimated (and for methods b and c deflated) for lower SRM levels. However, this reduction is somewhat greater for methods b and c. This variation reduces with increasing size of modelled SRM. The data are plotted in figure 3.3.7 (means) and figure 3.3.8 (standard deviations).

Table 3.3.5 Mean change scores and standard deviations for each computational method at each modelled SRM (using sample size of 90)

Modelled values		Method a		Method b		Method c	
SRM	SD	Mean	SD	Mean	Sd	Mean	Sd
0.1	10	0.9799	9.9636	8.0164	6.0261	8.3499	6.1533
0.25	4	0.9971	3.9952	3.2946	2.4783	3.5840	2.5899
0.5	2	0.9990	1.9917	1.7875	1.3335	2.0146	1.3869
0.8	1.25	1.0029	1.2455	1.3022	0.9298	1.4605	0.9395
1.0	1	1.0059	0.9981	1.1727	0.7968	1.2934	0.7907
1.25	0.8	1.0021	0.7979	1.0830	0.6848	1.1639	0.6691
1.6	0.625	0.9988	0.6223	1.0274	0.5742	1.0719	0.5563
2.0	0.5	1.0020	0.4973	1.0105	0.4801	1.0293	0.4684

Data drawn from main modelling programs

Figure 3.3.7 Comparison of calculated mean change scores for each method (using sample size of 90)

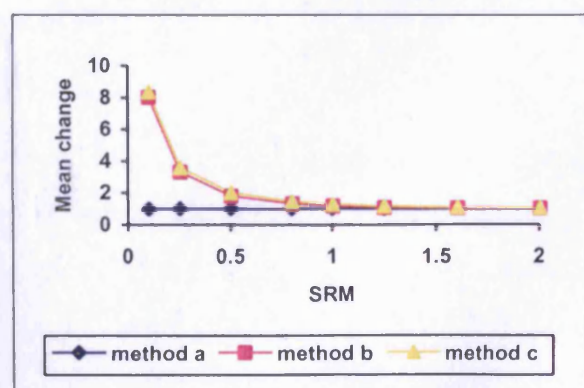
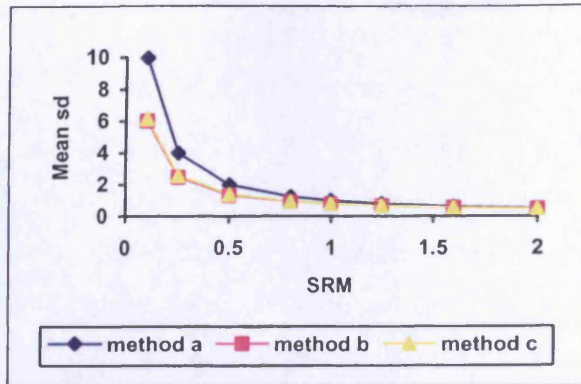


Figure 3.3.8 Comparison of calculated standard deviation scores for each method (using sample size of 90)



The three methods of dealing with negative values resulted in large differences in resultant pSRMs and npSRMs, especially within the range of effect sizes that are considered to be important.⁴⁴ Furthermore, only for method a was the value of SRM set by the modelling program equivalent to the 'true' SRM. Therefore, it was decided to restrict assessment of relative bias to method a.

Relative bias of derived SRMs: The bias in derived SRM compared to the modelled SRM was calculated by the modelling program for each SRM and sample size. This was conducted for both pSRMs and npSRMs. The relative bias values are shown in table 3.3.6. For the parametric models, generally the amount of bias was minimal and further, decreased with increasing sample size. The bias was always positive (i.e. overestimating the modelled SRM). With a sample size of 30 the relative bias was no greater than 2% except when modelling the smallest SRM (0.1). Using the smallest sample size (n=15), relative bias was under 7% in all of the modelled SRMs. For the largest sample size modelled, the relative bias was no greater than 0.3% except for SRM 0.1, where it was just over 0.5%. The general pattern of decreasing bias with increasing sample size was found for all SRMs modelled except for SRM 0.1. In this case, bias increased with a sample size of 30 and remained higher than for other SRMs at sample size 45 before returning to a level consistent with the other modelled SRMs. The distributions of relative bias values (for pSRMs) are shown in figure 3.3.9.

The pattern of relative bias values for npSRMs was similar to that of pSRMs (figures 3.3.9 and 3.3.10). In fact, derived npSRMs were slightly less biased than their parametric equivalents (table 3.3.6) with no value greater than 4.2% when using the smallest sample size. Again, relative bias generally decreased with increasing sample size and there was little difference in relative bias across different SRMs modelled. The bias was generally positive (again over-estimating the 'true' SRM) but there were a few instances of negative values in models of the smaller SRMs (0.5 or less). Again the smallest SRM, 0.1, displayed a slightly different pattern of relative bias values – similar to that seen for the pSRMs (figure 3.3.10).

Table 3.3.6 Relative bias (%) of pSRMs and npSRMs: by different SRM and sample size (method a – values drawn from a normal distribution)

	SRM	Sample size					
		15	30	45	90	180	360
pSRM	0.1	5.4462	7.3079	4.7302	1.0201	0.5619	0.5485
	0.25	6.1845	2.4158	1.2848	1.2065	0.3147	0.0700
	0.5	5.9686	2.8424	1.9330	0.9658	0.4124	0.1751
	0.8	5.5518	2.6751	1.6393	1.0114	0.4568	0.2078
	1.0	5.2595	2.4454	1.5159	1.0070	0.5733	0.2599
	1.25	6.6968	2.7458	1.3924	0.7172	0.5070	0.1485
	1.6	6.0116	2.6230	1.6473	0.8991	0.4750	0.1477
	2.0	6.1075	2.7961	1.6553	1.0639	0.3516	0.2214
npSRM	0.1	1.4341	5.9798	3.2895	-0.0210	0.2760	0.8127
	0.25	3.5203	0.9931	-0.0689	1.0018	0.4066	0.0453
	0.5	3.9713	1.6712	0.7312	0.2695	0.0615	-0.1569
	0.8	3.1107	1.1326	0.7021	0.5475	0.3645	0.1706
	1.0	3.3977	1.0472	0.5958	0.6842	0.3794	0.0943
	1.25	4.1087	1.1503	0.2924	0.4584	0.3236	0.0918
	1.6	3.5753	0.8819	0.6672	0.5360	0.2435	0.0740
	2.0	3.3201	1.1806	0.8173	0.5615	0.1625	0.1691

Relative bias values plotted in figures 3.3.9 (pSRM) and 3.3.10 (npSRM)

Figure 3.3.9 Relative bias plot for pSRMs by sample size and true SRM (method a – normal distribution)

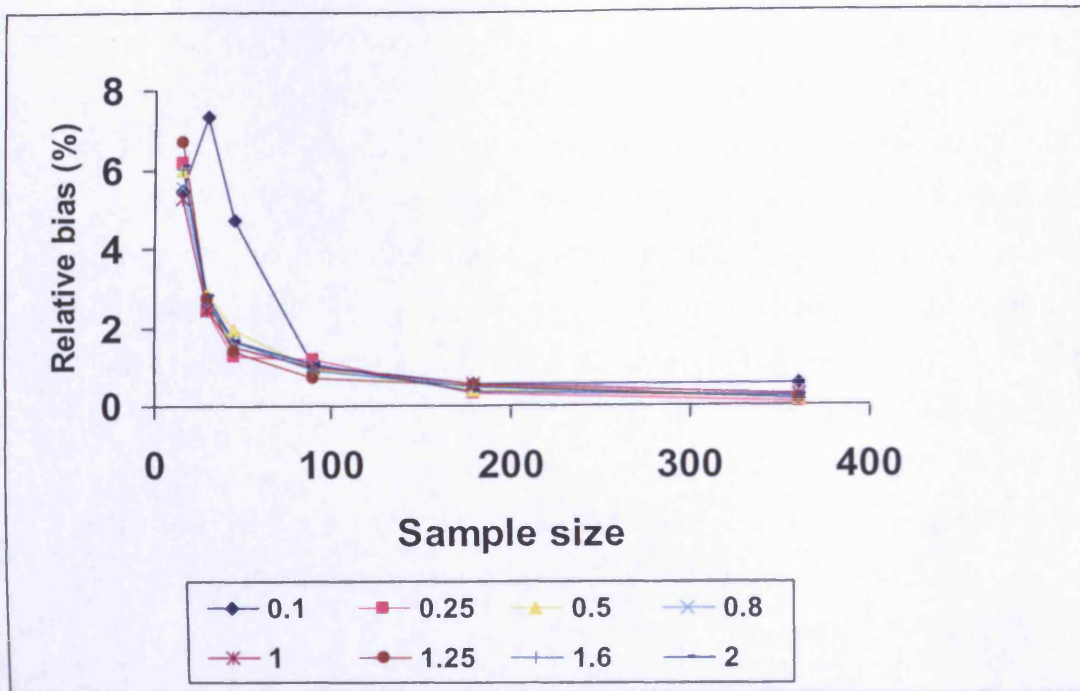
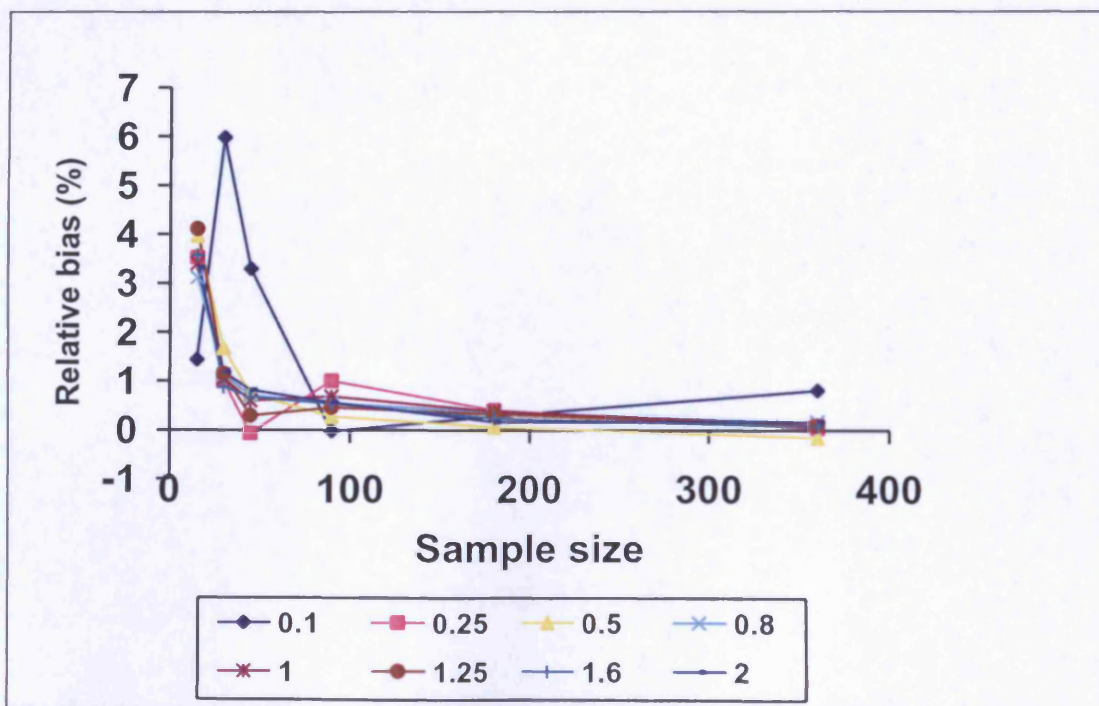


Figure 3.3.10 Relative bias plot for npSRMs by sample size and true SRM (method a – normal distribution)



3.4 Study II: Relative bias of SRM given a non-normal (skewed) distribution

Methods

In study I, values were generated by the modelling program based on parameters (mean and standard deviation) for an underlying normal distribution. Values generated in study II were to represent an underlying log normal distribution. Such a distribution is one that, when logged, approximates a normal distribution. In the modelling program, it was therefore, created by firstly, generating values from a normal distribution and secondly, taking the exponential of each value generated.

Comparable values of SRM were used in study II to study I (i.e. 0.1 to 2). The mean and standard deviation of the log normal distribution were, therefore, specified (as they had been for the normal distribution in study I). However, the required parameters of the initial normal distribution had to be identified and used in generating values. The specified log normal distribution values for mean and standard deviation were used to calculate the normal distribution values by applying formulae derived from Lindgren (figure 3.4.1).⁸⁴ In the modelling program, random numbers were thus generated using the derived normal distribution parameter values. Finally, the exponential of each generated value was taken.

Figure 3.4.1 Formulae used to derive normal mean and standard deviations from log normal parameters

$$1) \quad nmu = \log(lmu) - \frac{nsd^2}{2}$$

$$2) \quad nsd = \sqrt{\log\left[\frac{lsd^2}{lmu^2} + 1\right]}$$

Where: nsd = sd of normal distribution

nmu = mean of normal distribution

lsd = sd of specified log-normal distribution

lmu = mean of specified log-normal distribution

Calculating true npSRM with log normal distribution: The parametric version of the 'true' SRM was calculated by dividing mean change by standard deviation of change. Both of these latter two values were specified in the modelling program. A non-parametric version of the 'true' SRM was calculated by dividing median change by the inter-quartile range. Both of these latter two values were derived in the following way:

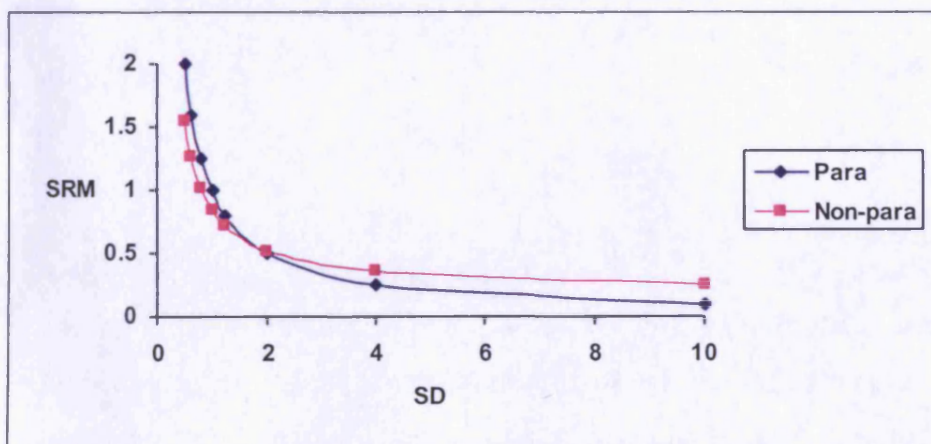
- 1) Required values (mean and standard deviation) for the log normal distribution to be modelled were specified
- 2) Equivalent values for (normal) distribution were calculated using the formulae in figure 3.4.1
- 3) Upper and lower quartiles for normal distribution parameters were calculated
- 4) Equivalent values for the required log normal distribution were derived by taking the exponential of mean, upper and lower quartiles
- 5) Non-parametric inter-quartile range and true npSRM were calculated

Determining bias: The modelling program calculated two estimates of SRM (parametric and non-parametric) for a single distribution of log normal values. Bias for each estimate was calculated relative to the respective 'true' SRM.

Thus, the estimated pSRM was compared with the 'true' pSRM (based upon set mean and standard deviation). The estimated npSRM was compared with the 'true' npSRM calculated by the modelling program.

Comparing true parametric and non-parametric SRMs: The true parametric and non-parametric SRMs for each set of specified mean change and standard deviation values (set parameters of the log normal distribution) are compared in figure 3.4.2. Both pSRMs and npSRMs decrease in value as variance increases. Below a standard deviation of 2, pSRMs have slightly greater value compared to non-parametric SRM. Above a standard deviation of 2, npSRMs are slightly larger than their parametric equivalent.

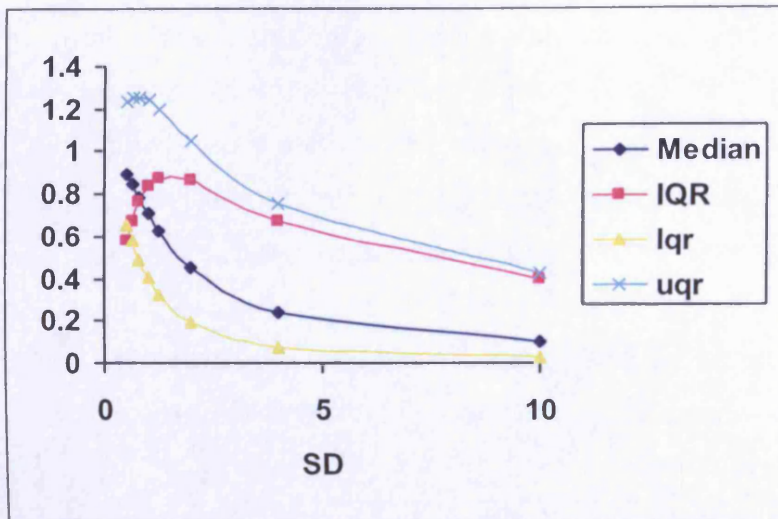
Figure 3.4.2 Comparison of true pSRM and npSRM for the range of standard deviations used in the modelling series – (log normal distribution)



To further assess the nature of the non-parametric SRMs associated with log normal data, the constituent medians, upper and lower quartiles used in calculating SRM are plotted together in figure 3.4.3. Whilst the values of both the lower and upper quartiles generally decrease as variance increases, for the latter there is an initial increase at smaller values of standard deviation. Consequently, the inter-quartile range increases as the standard deviation rises above 0.5 and starts to decrease when the standard deviation rises above 1.25. The median, like the lower quartile decreases in value steadily across the modelled range of standard deviations. Therefore, it is this

fluctuation in value of the upper quartile that appears to influence the shape of resultant non-parametric SRMs.

Figure 3.4.3 Component parameters used in deriving npSRM in simulated log normal distribution.¹



¹ Calculated by programme log2_28a.ftn.

Results

The relative bias of pSRMs and npSRMs for sample sizes 15 to 360 is shown in table 3.4.1 and plotted in figures 3.4.4 and 3.4.5. Estimated pSRMs show substantial positive bias (i.e. an overestimation) compared to the 'true' pSRM based upon set mean and standard deviation. Whilst the bias decreased with increasing size of SRM, in smaller samples it could still be substantial (for example, 11.5% bias; sample size = 15; SRM = 2). In models with a small SRM the bias was always substantial (for example, 142% in a sample of 360 and SRM of 0.1). Where the modelled SRM was by convention 'large' (greater than 0.8) bias could be great given a modest sample size (e.g. 23% bias; sample size = 30; SRM = 0.8).

Table 3.4.1 Relative bias (%) of pSRMs and npSRMs: by different SRM and sample size (values from log-normal distribution)

	SRM	Sample size					
		15	30	45	90	180	360
pSRM	0.1	441.2157	339.4283	294.0082	228.2031	179.3881	142.2164
	0.25	161.0229	121.3887	102.3711	76.4945	57.6648	43.7176
	0.5	67.5662	47.2600	38.4696	27.0730	18.2675	12.8429
	0.8	35.5165	23.0724	17.8230	11.7432	7.0653	4.4523
	1.0	26.6372	16.2693	12.1888	7.4912	4.5113	2.6578
	1.25	18.9557	11.6661	8.3496	4.7724	2.7918	1.4875
	1.6	14.3936	8.0484	5.4173	3.2742	1.7828	0.8251
	2.0	11.4739	6.2570	4.4780	2.3392	1.3645	0.5918
npSRM	0.1	17.5284	6.4712	5.2613	2.6648	1.7850	0.5488
	0.25	10.5161	3.8918	2.8435	1.8279	0.9882	0.2772
	0.5	7.6476	3.0274	1.7575	1.1678	0.6337	0.3011
	0.8	4.7433	2.3159	1.6579	0.5466	0.3090	0.2936
	1.0	5.1376	1.8198	1.3835	0.6418	0.2540	0.1875
	1.25	4.7701	1.5603	1.0933	0.8141	0.4633	0.2377
	1.6	3.9357	1.4291	0.8531	0.5770	0.2671	0.1878
	2.0	3.5653	1.8406	1.1125	0.5033	0.2861	0.2279

Relative bias values plotted in figures 3.4.4 (pSRM) and 3.4.5 (npSRM)

Bias in estimated npSRM relative to 'true' non-parametric SRM was much less than found for pSRM. The largest observed bias was 17% (in the smallest sample and SRM combination). Bias decreased with increased sample size and SRM.

Figure 3.4.4 Relative bias plot for pSRMs, by sample size and true SRM (log-normal distribution)

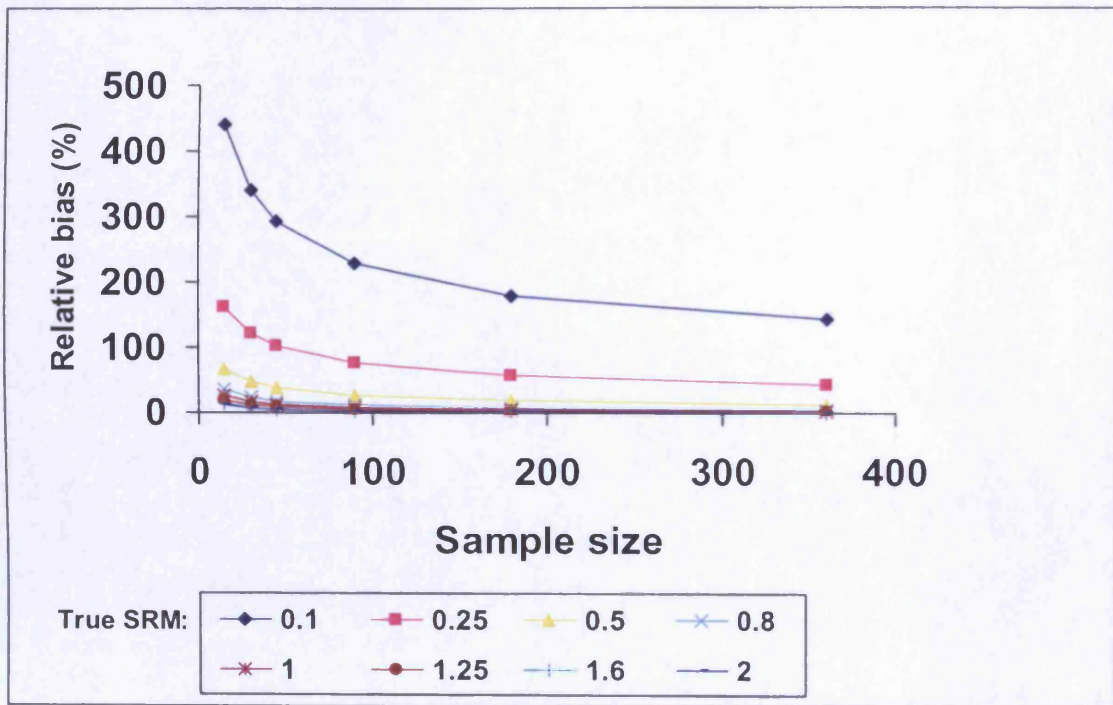
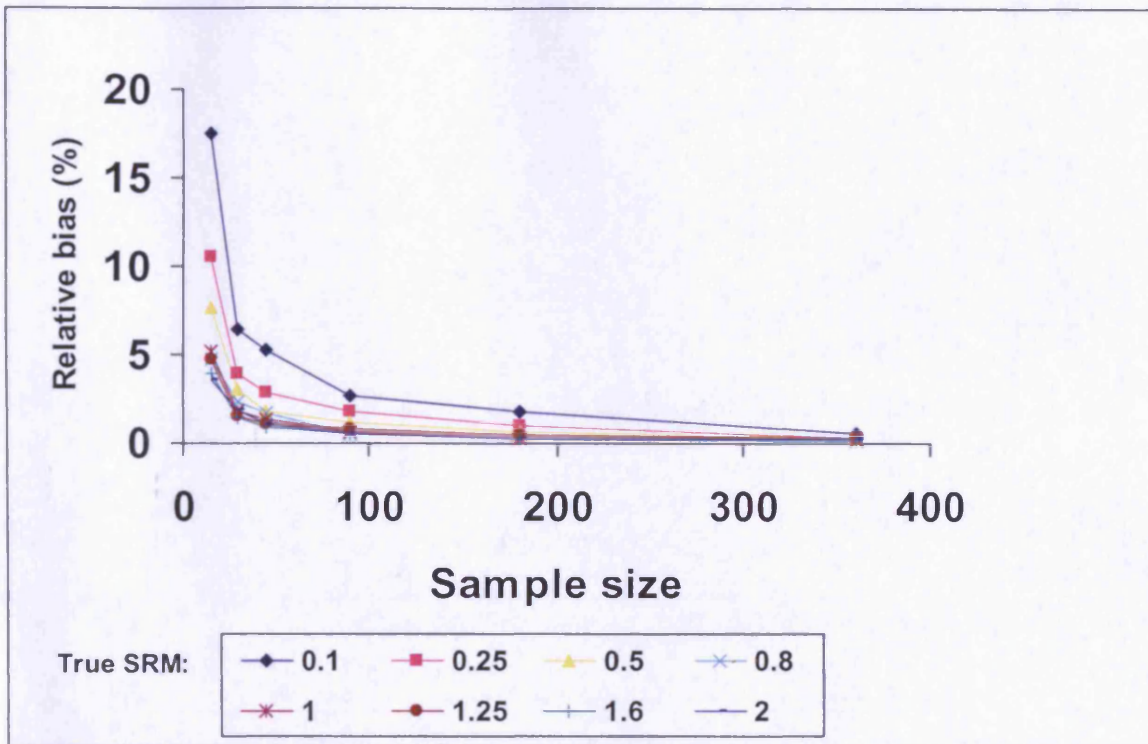
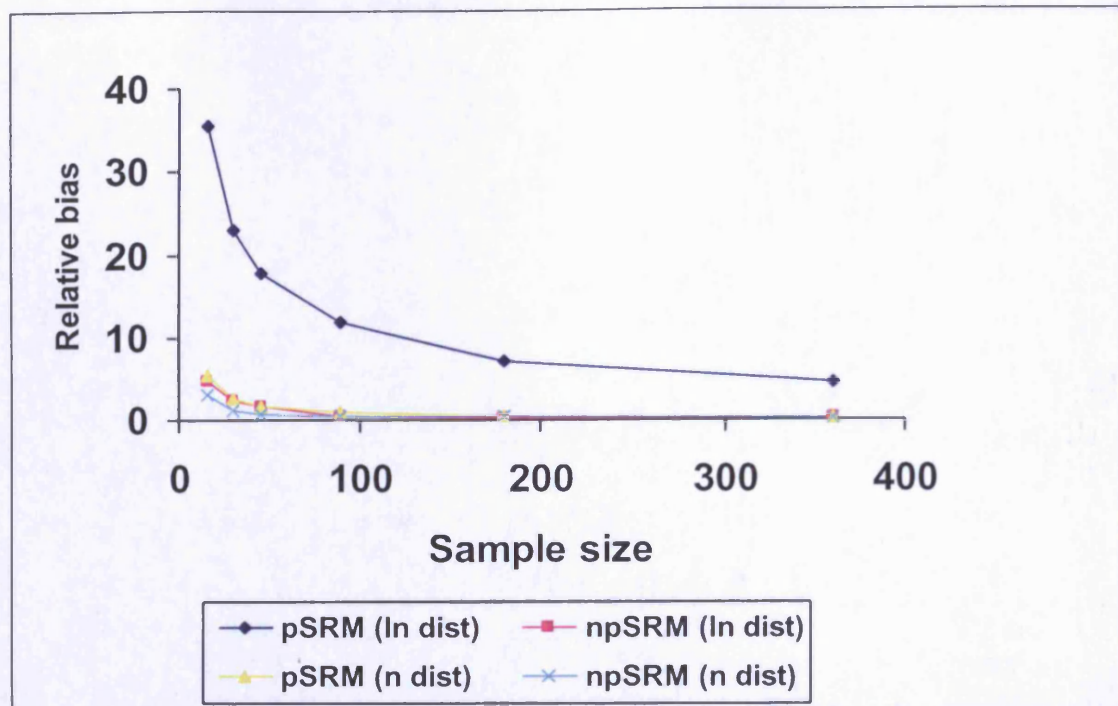


Figure 3.4.5 Relative bias plot for npSRMs, by sample size and true SRM (log-normal distribution)



Biases in estimated npSRMs and pSRMs are directly compared using data from both studies I and II in figure 3.4.6 (using data for a true SRM of 0.8). In all cases bias reduced with increasing sample size. Bias in both pSRM and npSRM was relatively small when data were drawn from a normal distribution. Similarly bias in npSRM was small when data were drawn from a log normal distribution. However, when a pSRM was estimated using data drawn from a log normal distribution the degree of bias was much greater.

Figure 3.4.6 Relative bias plots for pSRMs and npSRMs by sample size for true SRM of 0.8 (normal and log-normal distribution)



3.5 Discussion

Overview: The two simulation studies have evaluated the impact of different methods of calculating standardised response mean and varying distributional parameters upon estimated SRM. The first study demonstrated that SRMs estimated by reversing the sign of negative cases, or omitting them completely were similar to each other but substantially different from SRMs calculated using all cases as generated. This was the case especially within the range that effect sizes are often reported. SRMs estimated when using all cases showed only small bias relative to the expected ‘true’ SRM, especially in larger samples. Even for small samples the bias was never greater than 8%. The same pattern of minimal bias was found for both parametric and non-parametric versions of SRM. Bias was nearly always positive, slightly inflating the estimated SRM compared to the true value. However, the second study showed that when data was generated from a log normal (skewed) distribution, parametric SRM estimates were substantially biased, especially with smaller samples and small true SRMs.

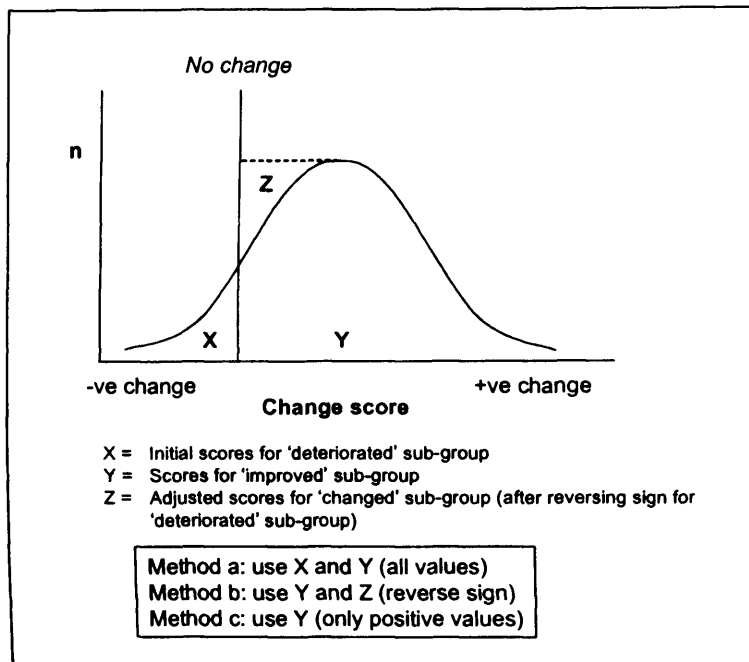
Methods of calculating SRM: Reversing the sign of negative values, or omitting them altogether, increases the mean change and reduces the standard deviation used when calculating the SRM (both consequences serving to increase the value of the subsequent SRM). This is particularly the case when the 'true' SRM is small (when there will be more negative values). When the 'true' SRM is large there is relatively little difference between the three approaches modelled. The simulation study demonstrates the important differences between the three approaches which can not be treated as equivalent. The true treatment effect for study patients who have either all received an intervention of proven efficacy, or only those self-reporting change (or simply improvement) is likely to be very different, even before consideration of the bias associated with an effect size statistic. Comparison of standardised response means from different studies should therefore, account for how the statistic has been calculated, and whether a like-for-like comparison is being made.

In study I, reversing negatively signed cases was used as a proxy for situations where a patient sub-group has been identified as deteriorated according to an external criterion of change (e.g. patient self-reported transition item). In practice, even when a transition item is used, subjects in the deteriorated sample may report positive change scores as well as negative. Similarly, the improved group may also contain negative change values. If the transition item is not sufficiently specific to the outcome measure of interest this may increase the likelihood of this. Furthermore, the reliability and validity of single-item transition scales for identifying stable and changed sub-groups has been questioned with concerns for example, about the influence of current health status upon rating.⁸⁵

Modelling a single population of change scores: Figure 3.5.1 represents the conceptualisation of the single change distribution used in the modelling study. Such a distribution may possibly be found in a single study group (i.e. the intervention sample). If negative cases (denoted by X) are omitted, or have their sign reversed (and therefore, increasing the number of positive

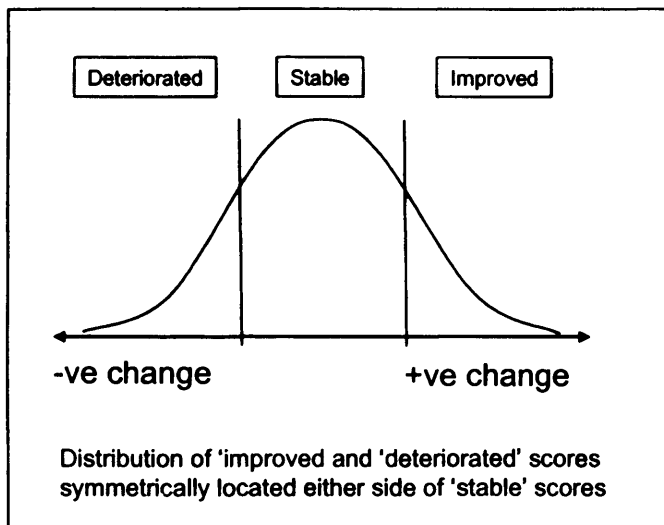
cases - denoted by z) one may see how this reduces variance in the resultant distribution and increases the mean change score. It will also affect the shape of the underlying distribution, tending to make it less normal.

Figure 3.5.1 Model of single change distribution



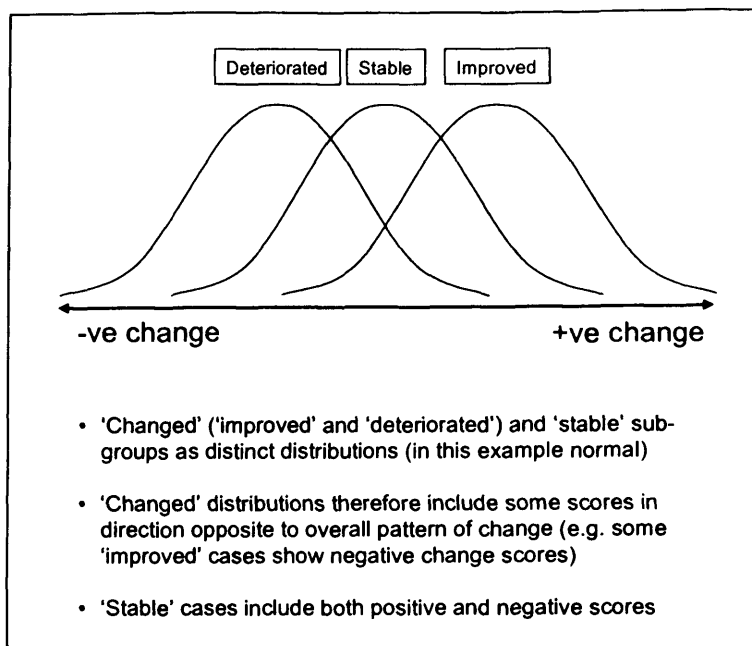
Norman and colleagues recommended confirming the equivalence of the means and standard deviations for improved and deteriorated sub-groups prior to combining to calculate an effect size.²⁷ Assuming a single distribution of change scores, this would require balance between improved and deteriorated patients (figure 3.5.2). Only when the change scores are symmetrically distributed around a mean change score of zero is this likely to be the case. Change scores for patients identified as stable could vary but would still have to average zero. Even when this is the case and improved and deteriorated cases are combined, the resultant distribution of change scores will not be normal, but may be highly skewed.

Figure 3.5.2 Model of balanced deteriorated and improved sub-groups



Multiple populations of change scores: An alternative model of change scores is one where patients identified as improved, stable or deteriorated represent discrete population sub-groups. Such a collection of distributions may possibly be found across study groups, for example, a control group being 'stable' (or 'deteriorated'), an intervention group being 'improved'. Alternatively, this may be the case for sub-groups of patients in a natural history study whose disease progression is determined by one or more varying risk factors. A model of multiple populations is represented graphically in figure 3.5.3. Reversing the sign of change scores for deteriorated patients would retain the shape of the resulting distribution. This would require though that both distributions of change scores for improved and deteriorated sub-groups were normal. However, the deteriorated and improved groups would still need to be similar in terms of mean change score and standard deviation. Furthermore, Cella and colleagues have questioned the clinical meaningfulness of combining improved and deteriorated change scores.⁸⁶ Whilst even small improvements in outcome measure score may represent important change for patients, possibly even large decreases in score may not. Lennert and colleagues have proposed response shift as a mechanism for explaining the discrepancies in patients' evaluation of improved or deteriorated health states.⁸⁷

Figure 3.5.3 Three distributions of change scores



A potential limitation of the current modelling study is the conceptualisation of a single change population to represent patients who change or remain stable. Simulating separate change distributions for each sub-group would be a useful alternative strategy to explore, especially if one wanted to model calculation of the index of responsiveness. The choice of mean change per sub-group and distributional equivalence of each sub-group would be key considerations in such a study. Nevertheless, the use of a single distribution of change scores to model the standardised response mean is more justifiable, especially if one is modelling the scenario of change following an intervention of known efficacy.

Relative bias: The degree of bias associated with parametric and non-parametric SRMs with an underlying normal distribution of change scores is small. Although the bias in smaller samples is greater, it is still no more than an 8% overestimate. When comparing parametric and non-parametric SRMs across different studies it is worth noting the different value of 'true' SRM produced by parametric and non-parametric methods. When data are drawn from a normal distribution the value of the non-parametric SRM is about 75% that of the parametric and should be taken into account when interpreting results.

The pattern of relative bias is consistent across all SRMs modelled in this study apart from the smallest (0.1). For both parametric and non-parametric SRMs this showed an increase in bias from small to moderate sample size (from 15 to 30/45) before reducing substantially. Derived non-parametric SRMs were less biased than parametric SRMs with an underlying normal distribution. The calculation of the non-parametric SRM may be less susceptible to influential outliers when small samples are used.

Studies with small sample sizes will inevitably result in a less precise estimate of responsiveness than larger studies. Although Liang and colleagues proposed the use of a 'jackknife' procedure for deriving confidence intervals for estimates of standardised response mean, there has been little evidence of this in practice.⁴⁰ However, more recently confidence intervals are being reported for a variety of effect size statistics (including standardised effect size and Index of Responsiveness) using methods such as bootstrapping.⁸⁸ Increased use should be made of confidence intervals to allow more informed inferences based on reported effect size estimates.

Log normal distributions: When data are drawn from an underlying skewed distribution (log normal), the non-parametric SRM again showed little bias in its estimate of 'true' SRM. Bias was slightly higher than that observed when a normal distribution was used to generate data, but mainly for small SRMs. Increasing the sample size reduced this bias even for smaller samples. However, relative bias was much greater for parametric SRMs calculated on log normal data. For a modestly sized sample ($n=30$) and a medium sized treatment effect (i.e. 0.5) this positive bias could be nearly 50%. The indiscriminate use of a parametric SRM with an underlying skewed distribution of change data could substantially alter the conclusion drawn from the use of an outcome measure. Whether or not change scores are initially normally distributed, the method chosen to study treatment effect may result in a skewed distribution (for example, selecting only 'improved' cases; selecting only those in the active intervention arm; combining improved and deteriorated sub-groups). The results of the present study emphasise the

need to apply an appropriately non-parametric SRM when data is skewed. Comparison of effect size data across studies should take skew of change scores into account.

Future directions: An advantage of modelling data using computer simulation is that a true value for the SRM can be precisely defined, and therefore bias in its estimation determined. The impact of changing various distributional parameters can also be precisely controlled and observed. However, some limitations apply. For example, the assumptions about what constitutes a changed or stable case. Further modelling could use separate change distributions, which may be individually varied. This may also be informed by further assessment of empirical data to address how varying cut-points for self-reported transition affects the constituency of sub-groups and consequently the size of observed effect. Similarity between distributional parameters for self-reported or study designed sub-groups would also be informative.

The simulation studies have sought to demonstrate how the validity of the SRM varies under different distributional conditions. However, the precise cause of the effects shown has still to be determined. Whilst identifying this lay outside the scope of this thesis, such an investigation (for example by further modelling or algebraic analysis) would be valuable.

3.6 Summary

The responsiveness of outcome measures is increasingly being reported using effect size statistics, with little apparent consideration of their performance under different distributional conditions and using different computational methods. These studies have demonstrated that:

- The method of calculating the standardised response mean has a substantial impact upon the SRM. In essence, the expected or true SRM differs according to the method chosen for including or excluding negatively signed cases. Comparison of such effect sizes across and

even within studies is potentially misleading if attention is not paid to method of calculation and study design. If an external criterion of change is not used to identify changed and stable sub-groups, calculation of SRM should therefore include all cases. When comparing the SRM for changed or stable sub-groups found in different studies or using different outcome measures there should be a consistent external criterion to ensure comparability of statistics. SRMs derived for general change scores (for example, in a single intervention group) and SRMs derived from sub-groups on the basis of self-reported health transition (e.g. the 'improved' sub-group in the same intervention group) should not be directly compared.

- Using all values as generated from an underlying normal distribution, the amount of bias in estimated SRM (relative to an underlying true value) was generally small. This was true even when varying the size and standard deviation of the underlying true change score. Although relative bias was larger in smaller samples, it was still no greater than 8% and provides re-assurance about the use of this statistic for reporting and comparing scale responsiveness.
- When change data were drawn from a log-normal distribution, a non-parametric version of the SRM (calculated from median and inter-quartile range) also showed little bias. However, the parametric SRM was substantially biased and provided a greatly inflated estimate of effect. This was especially so when the sample size and underlying SRM were small.
- Therefore, the less commonly used non-parametric SRM should be applied when reporting skewed data to avoid inflation of reported SRM and the risk of a Type I error.

The current chapter concludes the substantive presentation of the responsiveness theme. Summary points from both responsiveness and response shift themes are synthesised in the concluding chapter of the thesis (chapter 13). The following chapter therefore introduces the second major theme of the thesis – response shift. This is then followed by the design and empirical chapters of this second theme.

Chapter 4 Assessing change and response shift

Abstract: *This chapter describes the growing interest in response shift, and consequent theoretical developments. There is discussion of the processes that may underlie response shift. A literature review is presented which addresses empirical studies of response shift. A secondary aim of the review was describe the context and choice of evaluative method. The particular methodological challenges that response shift creates within evaluative research are highlighted. Response shift may pose difficulties in clinical settings but considering response shift may help clarify patient experience and inform the development of clinical interventions. Such possibilities are discussed. Finally, some of the approaches to detecting and exploring the nature of response shift are briefly reviewed. The chapter sets out the basis for the response shift themed studies of the thesis.*

4.1 The emergence and theoretical modelling of response shift

The measurement of a target construct (such as quality of life) on at least two occasions (usually before and after an intervention) is central to the assessment of scale responsiveness in most methods. It is assumed that observed changes in level of quality of life reflect the effects of the intervention. However, in longitudinal assessment other factors such as social desirability, effort justification and cognitive dissonance reduction may influence subjective self-report. Response shift is a key emerging concept in clinical assessment, and one that is also likely to influence reported change in quality of life.

Recognising different forms of change: Understanding what is being measured by change over time in subjective self-reported experience is central to the validity of much health science evaluation. That the interpretation of such changes observed in experimental and other longitudinal designs is more complex than that for physiological variables, for example, has been long recognised. In 1976, Golembiewski and colleagues

distinguished three relevant types of change in studies of organisational development - alpha, beta and gamma,⁸⁹ Only the first of these may be of primary interest to evaluation researchers. Alpha change involves variation in the level of an existential state given a constantly calibrated measure. In a clinical context examples of such a state may be pain or quality of life. Critically, measuring alpha change requires reference to a constant conceptual domain – the patient's understanding of the evaluated construct remains the same at each assessment.

Beta change, however, refers to the re-calibration of the interval used for measuring an otherwise stable conceptual domain. Respondents may expand or contract their own internal scale (at either or both ends of a theoretical continuum) for answering a survey question over time. Golembiewski describes this as a change in the psychological space between some intervals of a likert instrument.⁸⁹ The loss of a common metric across assessment time-points thus confounds simple pre-post comparison. Finally, gamma change represents a re-conceptualisation of the subjective domain of interest. This is a fundamental change in the frame of reference by which the construct of interest is evaluated by an individual. For example, how an individual understands their own quality of life may have changed between assessments.

Complementary work on re-calibration: At around the same time as Golembiewski's work emerged, studies by Howard and colleagues in the field of education and training also identified the occurrence of internal scale re-calibration (analogous to beta change).^{90 91} This work was partly stimulated by apparent discrepancies between negative research findings and subjective impressions of intervention benefits.⁹⁰ Regarded as a threat to internal validity, such changes were termed 'response-shifts' by Howard. Response shift was equated to the 'Instrumentation' bias identified by Campbell, in this instance the measuring instrument being the rater (e.g. patient) themselves.⁹⁰⁹² Thus, response shift was essentially viewed as deleterious, at least from a measurement perspective.⁹³ The potential for this bias was considered to be greater if an intervention sought to amend the subject's awareness or

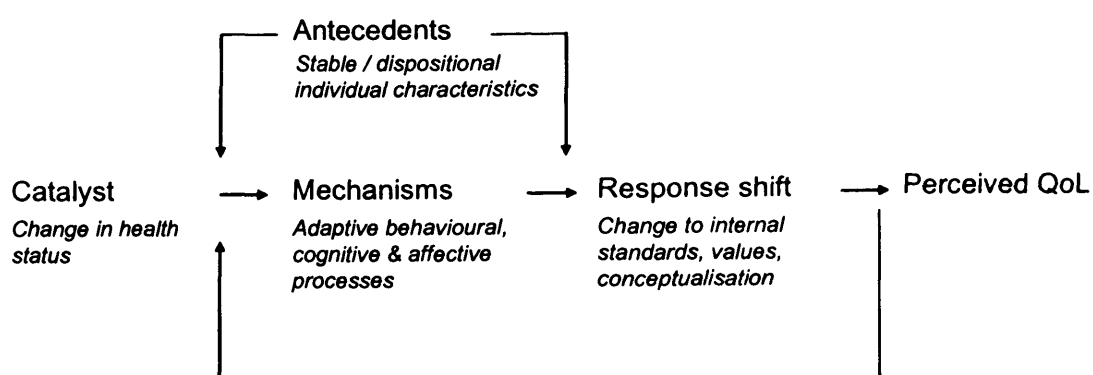
knowledge of the construct being measured (for example, an attitude). However, both Howard and Golembiewski note that producing a response shift may well be an intended aim of an intervention, especially if its purpose is to change subjects' understanding or awareness of a dimension under study.^{89 90}

Howard and colleagues described a cognitive process model of re-calibration whereby individuals form judgements about how their experiences relate to points on a response scale.⁹⁰ The questionnaire response scale represents a stable continuum (e.g. marks 1-10) which may be stretched or contracted to reflect an individual's experience.⁹⁰ An individual's awareness of the degree of possible experience may expand or contract at either end of the continuum. For example, a patient's experience of fatigue following cancer therapy may expand their awareness of the negative possibilities for fatigue. A state considered 'worst possible' prior to treatment (for example, coded 1), may subsequently be re-considered as further along the continuum (and coded, for example, 3).

An integrated model of response shift: More recently, Sprangers and Schwartz presented a theoretical model of response shift as it may affect health-related quality of life as a result of changes in health (figure 4.1.1).¹⁰ Described as a 'meta-construct' of three interrelated components, response shift was defined as "a change in the meaning of one's self-evaluation of a target construct due to change in the respondent's internal standards of measurement, change in values, or a re-definition of the target construct" (re-calibration, re-prioritisation and re-conceptualisation respectively).⁹⁴ Furthermore, the model incorporates the following elements: a catalyst; antecedents; mechanisms; response shift and perceived quality of life. Response shift changes are precipitated by a catalyst such as a change in health which in turn invokes an adaptive mechanism such as coping or goal reordering. Certain stable or dispositional individual characteristics such as personality (antecedents) may moderate how such mechanisms effect a change in internal standards, values or conceptualisation (response shift).

Finally, perceived quality of life alters as a consequence of the change in the individual's self-evaluation of that construct.

Figure 4.1.1 Theoretical model of response shift (adapted from Sprangers & Schwartz, 1999)¹⁰



Whilst Howard introduced re-calibration as the principal focus of response shift, Golembiewski additionally described re-conceptualisation.^{89 90} The above response shift model of Sprangers and Schwartz differed from earlier formulations by making explicit as a separate third component *re-prioritisation* (previously only implicit in Golembiewski's model of change).¹⁰ The response shift model was not intended to supplant existing theories of change or adaptation but rather to be integrated with them. It was considered that delineating the three separate sub-components of response shift would increase the likelihood of more fully capturing change.

An appraisal-based approach: Rapkin and Schwartz further developed this model of response shift in 2004 to produce a measurement model which specifically addressed the "phenomenology of qol appraisal".⁹⁵ As such, response shift is considered as a facet of a wider mechanism of individual self-appraisal. Schwartz and Sprangers noted within their original specification of the response shift model that it was not fully clear how to distinguish between components of the model.⁹⁶ For example, they described the similarity in how *mechanisms* and *response shift* were operationalised. The revised model was intended to clarify this distinction and also to differentiate

response shift (as an initial response to a catalyst) from feedback phenomena that served to continue the process.

Within the revised model response to quality of life items is regarded as contingent upon the appraisal process. Therefore, cognitive parameters related to coping and adjustment are delineated. The appraisal model may be viewed as analogous to those outlined in various cognitive models of survey response.⁹⁷ Common elements within such models include comprehension, retrieval, judgement and response generation.⁸ However, Rapkin and Schwartz considered the potential for their model's clinical application to mark it out as distinct.⁹⁷ Specifically, the appraisal parameters are 1) an induction of frame of reference; 2) the recall and sampling of salient experiences; 3) standards of comparison to appraise experience; and 4) a subjective algorithm to prioritize and combine appraisals to provide quality of life rating.

In demonstrating their model, Rapkin and Schwartz described response shift firstly within a regression paradigm and secondly, within a clinical perspective. Within the former, response shift was defined in terms of residual variance in change score explained by changes in appraisal (e.g. coping) once standard influences (i.e. catalysts) have been accounted for. Whilst this was not tested with real data, it does provide an evaluative model of response shift. The relationship between the various theoretical models of change, response shift components and appraisal parameters is shown in table 4.1.1.

Secondly, within a clinical paradigm, Rapkin and Schwartz described response shift in terms of the discrepancy between self-report and an external criterion (e.g. clinician judgement, performance test or caregiver assessment). Both judgements are considered subject to catalysts, antecedents and coping. Response shift would be implicated if a large proportion of the variance in the discrepancy between the two judgements was explained by changes in appraisal processes.

Table 4.1.1 Relationship between theoretical models of change and response shift

Golembiewski (1975) ⁸⁹	Howard and colleagues (1979) ⁹⁰	Sprangers & Schwartz (1999) ¹⁰	Rapkin & Schwartz (2004) ⁹⁵
Description of change typology	Response shift as a source of internal invalidity	Response shift as a meta-construct of three components	Appraisal framework for response shift
Alpha: change in existential level given constantly calibrated measure and stable conceptual domain			
Beta: re-calibration of measurement interval within stable conceptual domain	Measurement scale response continuum stretched to fit individual's experience continuum	Re-calibration: change in an individual's internal standards of measurement	Change in standards of comparison
Gamma: change in frame of reference of conceptual domain		Re-prioritisation: change in the importance of component domains constituting the target construct	Change in strategy for sampling experience within frame of reference relevant to quality of life rating Change in factors determining relative salience of different experiences
		Re-conceptualisation: re-definition of the target construct	Change in frame of reference

4.2 Mechanisms of response shift

Change in self-awareness and understanding: Sprangers and Schwartz suggested that adaptive mechanisms (behavioural, cognitive and affective) give rise to response shift following a change in health.¹⁰ In evaluating the impact of educational interventions upon outcomes such as attitudes or knowledge, response shift may occur because at baseline participants had only partially developed conceptualisations of the dimension being measured.⁹⁸ The intervention either increases awareness or understanding of this dimension, or an individual's insight into their level of functioning. This mainly affects re-calibration. In a clinical context this may still have validity for outcomes such as fatigue or pain if patients subsequently suffer levels of morbidity beyond their previous experience. However, for outcomes such as quality of life, response shift may not simply be due to acquisition of knowledge about the dimension as the domains that contribute to this subjective evaluation are re-configured (re-conceptualised).

Impression management: Sprangers and Hoogstraten experimentally eliminated response shift effects by convincing subjects that their self-ratings could be objectively validated.⁹⁹ From this they concluded that response style effects including social desirability bias and impression management were the cause of observed response shift. However, the recalibration response shift being assessed in their "thentest" study was in perceived ability. It may be that as an alternative explanation, social desirability may be more relevant in highly stigmatised therapeutic areas.¹⁰⁰ Re-prioritisation and recalibration response shifts within other subjective domains may be less likely due to the operation of such biases.

Social comparison: Another possible mechanism is social comparison, whereby changes in health status change the amount and type of comparison an individual engages in.^{101 102} For example, a negative change in health status may result in comparison with others at a lower level of functioning. The stress accompanying deteriorated health would also increase the amount

of comparison one engages in. Patients with multiple sclerosis, who participated in a coping skills intervention group, were thought to have maintained well-being despite physical deterioration, through a process of downward social comparison.¹⁰³ However, neither the evaluative process nor response shift were measured in this study by Schwartz. A second example by Van der Zee and colleagues found that selective downward comparison contributed to a sense of relative well-being amongst cancer patients.¹⁰²

A study of patients undergoing coronary artery surgery found that engagement in downward comparison was associated with better emotional and functional status.¹⁰⁴ Such social comparison may be viewed as adaptive, as it serves to maintain well-being. However, downward social comparison may also be maladaptive if it indicated a negative vision of the future for the individual which they regarded as inevitable.^{101 105} Furthermore, Gibbons suggested that longer-term downward comparison may be maladaptive if it leads to a deterioration in self-concept.¹⁰¹ 'Choosing' to either compare or contrast oneself with downward social comparators may effect whether the outcome is positive or negative for the individual.¹⁰¹

Downward social comparison may function within a buffering model whereby recalibration attenuates the association between health and quality of life.¹⁰⁶ Lepore and Eton evaluated a cognitive buffering model as a mechanism of response shift in patients with prostate cancer. In the face of physical deterioration, patients whose life goals changed reported improved quality of life. Patients whose goals remained stable reported worse quality of life. Whilst this was taken as evidence for a buffering effect, the authors did not find support for a suppressor model (where health problems facilitate a response shift, which in turn maintains perceived quality of life). Hagedoorn and colleagues also found evidence consistent with a buffering model of response shift in patients with cancer.¹⁰⁷ Those in physical decline who considered themselves to be better off than others, maintained better quality of life than those who felt the same or worse in comparison. However, Hagedoorn did not assess response shift directly and instead assumed that relative evaluations were the result of a social comparison.

Adaptive self-regulation: Carver and Scheier considered response shift within the context of adaptive self-regulatory systems of goal-seeking and affect-management.¹⁰⁸ Reference values within each system undergo gradual recalibration which may become apparent after periods of prolonged goal attainment or adversity. Discrepancy reducing feedback loops strive to make reality match a goal or standard of behaviour. Inability to attain goals may lead to a shift in standards or ultimately a disengagement from the goal. One goal may be traded for a different goal, although the new goal may still serve the same higher order function and thus the process may not be indiscriminate. Carver and Scheier equate the resetting of reference values to re-calibration, whilst goal substitution provides a mechanism for re-prioritisation and re-conceptualisation.

Response shift and coping: Richards and colleagues related response shift to a revision theory of coping in which thoughts and behaviours may be either emotion-focused (used to regulate distress); problem-focused (to manage problems causing distress); or meaning-focused (to maintain positive well-being).¹⁰⁹ In their view, response shift results from specific meaning-based coping processes that operate when previous beliefs, expectations and goals are no longer tenable. The author's conceptualisation is consistent with the Sprangers and Schwartz model of coping as one mechanism supporting response shift.

Permanence of change: There is debate as to whether response shift represents a temporary or permanent change in state.⁹³ For example, Norman and colleagues described response shift as a unidirectional and permanent change, whilst short-term fluctuations simply represent noise.^{93 110} In contrast, Sprangers and Schwartz's theoretical formulation, with the maintenance or regaining of homeostasis as its goal, emphasises dynamism and flexibility.¹⁰ Brossart argued for maturation as an alternative explanation for response shift if changes are expected given the age and developmental stage of the individual.¹¹¹ However, this may be less clearly defined in adults compared to children. Nevertheless, whether caused by developmental factors or

maturation, or as a response to a life event such as a change in health status, the net effect may be the same – re-calibration or re-conceptualisation of target construct.

Brossart and colleagues were particularly interested in the timing of a response shift and usefully considered two types of change – a one-time transient increase or decrease and a permanent change in level.¹¹¹ Within a growth modelling paradigm they considered other patterns of response shift. Perhaps linking the temporary and permanent divide is a view that response shift is essentially an iterative process of successive approximations.

Schwartz and colleagues discussed the results of a peer support intervention which, incidentally, focused participants upon the concerns of others.¹¹² The authors described how response shift may have occurred with individuals initially disengaging from fixed patterns of self-reference. Individuals were consequently more open to changes to internal standards, values and conceptualisation, from which followed a re-integration of the health problem and its personal meaning for the individual.

Catalysts for change: Whether response shift is necessarily precipitated by disease progress and treatment impact, or whether it may emerge simply over time, is also contentious.⁹³ Howard and colleagues originally noted that response shift effects were larger in treatment groups within experimental studies, thereby supporting the notion that response shift is treatment dependent.⁹⁰ Golembiewski also viewed gamma change (re-conceptualisation) as consequent upon an intervention.⁸⁹ However, others have argued that, in theory, re-conceptualisation may occur in either treatment groups, control groups or in both.¹¹³ Thus, Millsap and Hartog (who operationally defined gamma change / re-conceptualisation as change in factorial structure of measured construct) described how differences in posttest factor matrices between study groups (differential change) may be taken as evidence of an intervention effect. Parallel change may also occur whereby factor changes occur similarly in both study groups, providing no evidence of an intervention effect. Whilst in Sprangers and Schwartz's model, response shift specified a catalyst (i.e. a change in health status), the authors

considered that change may also occur with the passage of time and without overt change in health. Related to this issue is whether non-health factors may influence response shifts or whether only health factors are relevant.⁹³ Whether response shift may occur solely in experimental study groups or additionally amongst control patients is a key distinction for the interpretation of trial results.

4.3 A review of empirical response shift studies

Aim

For this thesis a literature search was conducted to determine the scope of response shift studies conducted in clinical populations and the range of methods used. The original special issue of *Social Science and Medicine* referred to above contained no papers where the assessment of response shift was the primary focus.¹¹⁴ The current review sought to identify the range of empirical studies where assessing response shift was either a primary or secondary analytic aim.

Search method

The search included the Ovid Medline, Embase and PsycInfo bibliographic databases (from 1966, 1980 and 1985 respectively, until Jan 2005) using the search term 'response shift' to identify relevant articles reporting empirical studies. Full articles not available in English and research dissertations were excluded from consideration. Non-empirical background articles were also identified. These included reviews, discussion papers, methodological papers and commentaries where response shift may have been either the focus of the paper or of secondary importance. All such papers were reviewed and listed in appendix III. The response shift phenomenon has been reported anecdotally; has been used subsequently to explain paradoxical findings (e.g. discrepancies between objective indicators and subjective accounts); and response shift effects have been described indirectly (for example, in descriptions of coping and adaptation). However, this review addressed response shift as a specifically defined and labelled phenomenon.

Of particular interest were the type of outcome being assessed (for example, quality of life) and the context of the study (e.g. educational, clinical or other). For clinical studies only (excluding, therefore, educational studies), the relevant patient or disease group was identified. The broad analytic method chosen was delineated, with classification based upon that described by Schwartz and Sprangers.¹¹ Where possible, the form of response shift addressed by the study was described according the theoretical model of Sprangers and Schwartz (i.e. change in internal standards, values or conceptualisation).¹⁰ For each study, whether or not the response shift analysis was a primary or secondary analytic aim at the time of original data collection was determined where possible. Furthermore whether data collection was longitudinal or cross-sectional was determined. Where relevant, the length of the test-retest interval was noted.

Findings

A total of 255 abstracts were identified. After removing non-relevant and duplicate abstracts, 91 relevant articles remained. Five dissertations and four non-English articles were excluded from further consideration. 52 relevant empirical (table 4.3.1) and 30 non-empirical papers (appendix III) were identified.

Table 4.3.1 Empirical studies related to response shift

Year	First author	Outcome	Context	Patient group	Analysis	RS type	Outcome data collection	Analytic aim	Re-test interval	Paper
2004	Ahmed	HRQoL	Clinical	Stroke	Design (thentest)	IS	Longitudinal	Primary	E: 6&24/52 C: 8/52	115
2004	Bernhard	QoL	Clinical	Cancer	Statistical Design (thentest)	C V	Longitudinal	Primary	Before / after surgery 2/12	116
2004	Chin	Sleepiness	Clinical	Obstructive sleep apnea-hypopnea	Design (thentest)	IS	Longitudinal	Primary	E: 10/12 C: 50 days	117
2004	Donovan	Fatigue	Clinical	Cancer	Statistical / Design	IS	Longitudinal	Primary	Before / during / after therapy	118
2004	Holzner	QoL	Clinical	Cancer	-	-	-	Discussion	Not applicable	119
2004	Kidd	Awareness of safety climate	Occupational health	-	Design (thentest)	IS	Longitudinal	Primary	4/12	120
2004	Paterson	Health status, QoL	Clinical	Various chronic illness – in receipt of acupuncture	Qualitative	IS, C	Longitudinal	Secondary	6/12	121
2004	Roos	QoL	Clinical	Gynaecological and bladder cancer	-	-	-	Discussion	-	122
2004	Schwartz	QoL	Clinical	Multiple sclerosis	Design (thentest), Statistical	IS C	Longitudinal	Primary	5 years	94
2004	Schwartz	Advance care planning preferences	Clinical	Various seriously ill (stable and declining)	Statistical	V	Longitudinal	Primary	21 days	123
2003	Essink-Bot	Health status, HRQoL	Clinical	Prostate cancer	-	-	-	Discussion	-	124
2003	Rees	LUT symptoms	Clinical	Cancer	Design (thentest)	IS	Longitudinal	Primary	3 & 6/12	125
2003	Schwartz	Physical & mental health	Community	-	-	-	-	Discussion	-	126
2003	Timmerman	QoL	Clinical	Children with OME	Design (thentest)	IS	Longitudinal	Primary	6-8/52	127

Year	First author	Outcome	Context	Patient group	Analysis	RS type	Outcome data collection	Analytic aim	Re-test interval	Paper
2002	Cella	HRQoL	Clinical	Cancer (various)	-	-	-	Discussion	-	86
2002	Greiffenstein	Scholastic performance	Legal / educational	-	-	-	-	Discussion	-	128
2002	Hagedoorn	Emotional functioning, global QoL	Clinical	Cancer	Statistical	IS	Longitudinal	Primary	3/12	107
2002	Joore	General & specific QoL	Clinical	Hearing impaired	Design (thentest)	IS	Longitudinal	Primary	12/52	129
2001	Arrindell	Psychiatric symptoms, state & trait	Clinical	Psychiatric in- & out-patients	-	-	-	Discussion	-	130
2001	Bernhard	Health utility	Clinical	Cancer	Design (thentest)	IS	Longitudinal	Primary	12 days 50 days	131
2001	Rohs	Behaviour	Educational	-	Design (thentest)	IS	Longitudinal	Primary	2/52	132
2000	Bar-On	QoL	Clinical	Hypertension	Statistical / design	IS, V	Longitudinal	Secondary	12/12	133
2000	Jansen	Fatigue, QoL, well-being	Clinical	Cancer	Design (thentest) Preference-based	IS V	Longitudinal	Primary	6-7/52	134
2000	Lenert	Health utility	Clinical	Primary care patients	-	-	-	Discussion	-	87
2000	Lepore	QoL	Clinical	Prostate cancer	Design (thentest) Qualitative	IS V	Longitudinal	Primary	10/52	106
2000	O'Boyle	QoL	Clinical	Cancer, elderly	Individualised	V, C	Longitudinal	Secondary	2, 12, 24/12	135
2000	Postulart	HRQoL, cost-effectiveness	Clinical	Diabetes, ESRD	Design (thentest)	IS	Longitudinal	Secondary	5, 12, 18/12	136
2000	Pratt	Parenting skills, knowledge	Social	-	Design (thentest)	IS	Longitudinal	Primary	6/12	137
2000	Rapkin	Global & emotional well-being, pain	Clinical	AIDS	Statistical	IS, C	Longitudinal	Primary	6/12	138
2000	Sprangers	Fatigue	Clinical	Cancer	Design (thentest)	IS	Longitudinal	Primary	4-7/52	139
2000	Thompson-Fawcett	QoL	Clinical	Cancer	-	-	-	Discussion	-	140
2000	Visser	Fatigue	Clinical	Cancer	Design (thentest)	IS	Longitudinal	Primary	Before / 2/52 / after therapy	141

Year	First author	Outcome	Context	Patient group	Analysis	RS type	Outcome data collection	Analytic aim	Re-test interval	Paper
1999	Daltroy	Physical function	Community	-	Statistical	IS	Cross-sectional	Primary	Not applicable	142
1999	Schwartz	QoL	Clinical	Cancer survivors	Design (thentest) Statistical	IS C	Longitudinal	Primary	Weekend 3/12	143
1999	Schwartz	QoL, well-being	Clinical	Multiple sclerosis	-	-	-	Discussion	-	112
1999	Schwartz	QoL, health status, well-being	Clinical	Multiple sclerosis	-	-	-	Discussion	-	103
1999	Sprangers	Fatigue	Clinical	Cancer	Design (thentest) Qualitative	IS	Longitudinal	Primary	4-7/52	144
1997	Manthei	Counselling skills / effectiveness	Educational / training	-	Design (thentest) Qualitative	IS	Longitudinal	Primary	7/12	145
1996	Hickok	Fatigue	Clinical	Lung cancer	-	-	-	Discussion	-	146
1996	Sprangers	QoL	Clinical	Cancer	Design (thentest)	IS	Longitudinal	Primary	Before / after therapy	147
1994	Robinson	Knowledge / skills	Educational / training	-	Design (thentest)	IS	Longitudinal	Secondary	Not described	100
1992	Goedhart	Management & leadership skills	Training	-	Design (thentest)	IS	Longitudinal	Primary	4/52	148
1992	Guttek	Job satisfaction	Employment	-	Design (thentest)	IS	S1-2: Cross-sectional / S3: Longitudinal	Secondary	18-20/12	149
1992	Skeff	Teaching performance / attitudes	Educational / training	-	Design (thentest)	IS	Longitudinal	Primary	Before, during & after course	98
1990	Levinson	Interviewing & teaching skills	Educational / training	-	Design (thentest)	IS	Longitudinal	Primary	5/52	150
1989	Sprangers	Communication skills	Educational / training	-	Design (thentest)	IS	Longitudinal	Primary	S1: 6/52 S2: 3/365	151
1988	Hoogstraten	Not specified	Educational / training	-	Design (thentest)	IS	Longitudinal	Secondary	Not described	152
1988	Sprangers	First aid skills / knowledge	Educational / training	-	Design (thentest)	IS	Longitudinal	Primary	25 mins	153

Year	First author	Outcome	Context	Patient group	Analysis	RS type	Outcome data collection	Analytic aim	Re-test interval	Paper
1987	Sprangers	Problem-solving	Educational / training	-	Design (thentest)	IS	Longitudinal	Primary	1 hour	⁹⁹
1987	Zwiebel	Attitudes to disabled people	Educational	-	Design (thentest)	IS	Longitudinal	Primary	Before / after course	¹⁵⁴
1985	Collins	Substance use self-report	Educational	-	Design	IS	Longitudinal	Primary	1-2.5yrs	¹⁵⁵
1985	Hoogstraten	Knowledge / functioning	Educational	-	Design (thentest)	IS	Longitudinal	Primary	S1: 25 mins S2: 45 mins	¹⁵⁶

For classification notes for table 4.3.1 see next page

Classification notes for table 4.3.1

Attribute	Level	Definition / notes
Outcome	Various	The main outcome assessed in the study: includes Quality of Life (e.g. general, specific, health-related), health utility, functional and health status, symptoms, knowledge, skills, behaviour, attitudes
Patient group	Various	For studies within a clinical context only - description of the main study population (may also include control group)
Analytic aim	Primary Secondary Discussion	The assessment of response shift conducted was an intended aim of the study's original data collection The assessment of response shift conducted was a secondary analysis of a study's originally collected data Response shift only considered in discussion (most often as an explanation for observations) within empirical study
Data collection ^a	Longitudinal Cross-sectional	Published study mainly presented data collected longitudinally Data collection was mainly collected in cross-section
Analysis ^b	Individualised Preference-based Successive comparison Design approach Statistical Qualitative	Use individually defined Quality of Life domains (e.g. Repertory Grid, SEIQoL) Importance and value accorded to health state (e.g. Q-TWIST method, preference mapping) Judgements of the ordering of attributes along a continuum (e.g. card sort approach) The study design is configured to determine response shift (e.g. thentest) Quantitative analysis of (primarily longitudinal) data (e.g. factor analysis, growth curve analysis) Qualitative single / group interviews
RS type ^c	IS V C	Change in internal standards Change in values Change in conceptualisation
Re-test interval	Various	Time period between assessments (for primary response shift studies)

a For studies only where primary aim of analysis was response shift

b Follows methodological framework of Schwartz and Sprangers¹¹

c Where specified

Study sample: A small number of studies were identified in non-clinical samples: education / training (n=14), social / community (n=3) employment (n=1), occupational health (n=1). However, the largest sub-group was of clinical (patient) studies (n=33). The roots of response shift research is reflected in the initial predominance of education and training studies identified and that clinical studies only started appearing in the mid-1990s. Within the clinical studies, the majority involved patients with cancer (n=19) with a variety of other conditions or patient groups also being represented (e.g. stroke, psychiatric in-and out-patients, multiple sclerosis). Most clinical samples thus involved serious, life-threatening, chronic or degenerative conditions. The most common outcome assessed was quality of life (including health-related, general and specific), which was reported in 21 studies. Other outcomes included fatigue (n=6 studies) reflecting the clinical focus upon cancer, and pain.

Methods used: The most common approach found was the design method, namely the retrospective pretest-posttest (thentest), found in 34 studies. A few of these studies compared only prospective and retrospective pretest ratings. An example of this was provided by Collins and colleagues, who assessed level of agreement between ratings.¹⁵⁵ The exclusive initial focus on design methods eventually expanded with greater subsequent diversity, for example, with the introduction of statistical and qualitative approaches. A small number of studies have integrated different methods to assess response shift, either to allow study of the different components, or the same component using a triangulating approach. Bernhard and colleagues present an example of the former, in their study of re-conceptualisation and re-prioritisation using statistical and design methods.¹¹⁶ Sprangers and colleagues describe an example of the latter when assessing re-calibration using a design method supplemented by qualitative interviews.¹⁴⁴

Focus of empirical studies: Re-calibration was the most commonly assessed component of response shift (n=37 studies), reflecting the dominant methodological approach (design). This probably also resulted from an early

theoretical view of response shift as primarily internal scale re-calibration (notwithstanding Golembiewski's typology including re-conceptualisation). Outcome data collection was predominantly longitudinal, with only two studies featuring the use of cross-sectional data.^{142 149} Whilst assessing response shift was the primary aim of data collection in 33 studies (n=19 in clinical studies), in several empirical studies this was not so. Some studies involved secondary data analysis, for example, Postular and colleagues studied response shift in diabetic patients with end-stage renal disease using a data set from a previous study.¹³⁶ In other studies response shift was an observed finding, rather than the primary purpose of the study (e.g. Paterson).¹²¹ Finally, response shift was provided as an explanation for observed findings in a number of studies which were not principally designed to directly address it (examples include Thompson-Fawcett and colleagues; Arrindell and colleagues; Hickok and colleagues).^{130 140 146}

Follow-up interval: The test-retest interval was noted where detailed, although for some studies reports only described assessment 'before and after' intervention (for example, Donovan and colleagues; Zweibel).^{118 154} The shortest reported interval of 25 minutes was reported in two educational experiments by Sprangers and Hoogstraten; and by Hoogstraten.^{153 156} Within clinical samples, short intervals were found in a number of studies, including a median of 12 days (range 6-73 days) in a study by Bernhard and colleagues assessing health utility in patients in receipt of surgical resection for adenocarcinoma of the colon.¹³¹ Another example of a short re-test interval (21 days) was a study of advanced care planning in seriously ill patients by Schwartz and colleagues.¹²³ Test-retest interval in the latter study reflected partly the aim of assessing outcome measure reliability and the serious nature of the patients' condition. The longest interval between assessments was reported by Schwartz and colleagues in a thentest study of patients with multiple sclerosis.⁹⁴ Respondents were followed-up after five years (range 50-69 months).

Background articles: A small number of the non-empirical articles identified also included some data analysis, but this was considered secondary to the

main review or, for example, used to illustrate a method. An example of the latter was a methodological paper by Lowy and Bernhard which demonstrated a multilevel modelling approach to assessing re-conceptualisation tested on a sample data set.¹⁵⁷ Such an approach addresses change in regression coefficients for component quality of life domains over time whilst adjusting for the repeated-measures assessment of individuals. Several of the articles were drawn from the Social Science and Medicine special issue on response shift or from the book 'Adaptation to Changing Health' by Schwartz and Sprangers.⁹ A small number of these and other papers, presented primary theoretical perspectives upon response shift (for example, Rapkin and Schwartz).⁹⁵ Some articles present response shift as one of several methodological considerations in quality of life or utility assessment, either in general or within a specific patient population (e.g. paediatrics or cancer).¹⁵⁸⁻¹⁶⁰ Further articles represent commentaries which evaluate empirical findings from the additional perspective of response shift.^{57 109}

Summary of literature review

The review has mapped the increasing interest and activity in studying response shift, especially within a clinical context. Amongst patient-based studies, cancer was the most commonly assessed condition but other mainly serious conditions have also been evaluated. The range and development of methods for assessing response shift is apparent, although the most commonly reported approach was the design method (specifically the "thentest"). Accordingly, re-calibration, which is assessed by the thentest, was the most commonly assessed response shift component.

4.4 Psychometric and clinical importance of response shift

The increasing importance of response shift was highlighted by a themed issue of Social Science and Medicine in 1999, and the subsequent publication of the book 'Adaptation to Changing Health' by Schwartz and Sprangers in 2000.^{9 114} Early descriptions of response shift firmly regarded it an 'unwanted' measurement bias that should be at least accounted for in study designs; controlled for; or even designed out of influence.^{90 91 100} The latter included

approaches such as the 'bogus pipeline' whereby respondents providing self-reports were misled into believing that self-reports could be objectively validated.^{99 153} Some more recent commentaries refer to the fundamental threat to the validity of self-report outcome data posed by response shift.^{111 114} However, this is accompanied by an awareness of its apparent ubiquity in subjective self-report and an increasing interest in its study, for example, as part of the normal process of adaptation.

Such recognition has led to a call to re-assess the role of psychometrics in the light of response shift.⁶⁵ Thus, Schwartz and Rapkin, called for a focus upon appraisal processes involved in responding to quality of life items - currently not widely addressed. Psychometric equivalence within this model would mean equivalent measures eliciting similar appraisal processes. An underlying true score would be contingent upon the process of quality of life appraisal, so that if appraisal mechanisms change the underlying true score also changes. An appraisal based approach would seek to distinguish between those who feel better from those who have changed their mind about what it means to feel bad.

Self-reports of subjective states: Self-report in educational, clinical and other settings offers advantages over, for example, the use of external raters, and may be the only realistic option for assessing essentially subjective outcomes (e.g. fatigue).¹⁴¹ Compared to the use of external raters (e.g. clinicians) self-reports are usually easier and more cost-effective to obtain.⁹⁸ However, response shift may potentially occur in any context - whether formal research or clinical care - where verbal or written self-report is required.^{111 161} Alongside other response mechanisms that influence self-reports, response shift effects require attention to the validity of such data. Wilson noted that as biological and physiological variables are not the product of self-report they can not be subject to response shift (although physical, rather than psychological, re-calibration may be a problem).¹⁶¹ However, evaluative clinical outcomes such as quality of life, which will have a large number of determinants, may be particularly susceptible to response shift.

The possibility of response shift demands closer attention to the meaning of self-report scores, and ultimately their validity. Unexpected results and discrepant observations raise the probability of multiple interpretations of change data and their interpretability.¹⁶² Expected differences in quality of life between patients in receipt of different treatment modalities for cancer; and between patients and non-patients have not always been found.¹⁶³ For example, Schwartz and colleagues noticed that a psychosocial intervention for survivors of childhood cancer apparently had an unexpectedly deleterious impact upon global quality of life.¹⁴³ Controlling the results for re-calibration response shift changed the direction of the apparent change. Cella and colleagues found sizeable improvements in quality of life over time for patients with advanced cancer, including patients providing baseline scores at the ceiling of the outcome scale.⁸⁶ Finally, Bernhard and colleagues found that following surgery, retrospective ratings of baseline quality of life were lower than those recorded prospectively in patients undergoing radical resection for colon cancer.¹⁶⁴

Observed treatment effect: Response shift may, therefore, change the expected direction of intervention effect, or it may simply reduce or inflate the observed effect size (a more subtle effect). Whilst this point is described in more detail in chapter 9, a recent review of 22 studies concluded a median effect size of +0.25 due to response shift.⁹³ Bray and colleagues suggested that the potential loss in statistical power due to response shift when unaccounted for in analysis could range from 5% to 90% (depending upon size of response shift and other parameters).¹⁶⁵ However, it should be noted that their estimates were derived from simulation studies using parameters informed by mostly educational research. Thus, Howard and colleagues concluded from their earlier studies that response shift would have led to an erroneous rejection of the experimental hypothesis.⁹⁰ Studies not accounting for response shift could face a loss of statistical power, a greater chance of a type II error and the rejection of potentially useful interventions.^{100 145} However, as Schwartz and colleagues noted in their review, response shift may both inflate or deflate the treatment effect.^{93 166}

Response shift may invalidate serial comparisons, but even true experimental designs may be affected.¹⁶⁶ Whilst Cronbach and Furby reject the use of change scores for analysing experimental studies in favour of comparison of only post-intervention test scores, even this may be invalidated with self-reported data.^{167 168} Bernhard and colleagues found a general reframing (recalibration) effect across treatment groups in a study of patients following surgery for colon cancer.¹⁶⁴ They suggested that reframing in their study may reflect a broad adaptation effect rather than being treatment specific. However, for one outcome (appetite) the effect still differed between study groups.

Clinical importance of response shift: Several authors have pointed to the clinical importance and application of response shift. Studying response shift may help map out the process of adapting to deteriorating health.⁹⁴ Similarly, explicit consideration of response shift may enhance study of the nature of quality of life experience.^{95 169} Response shift may result in under-reporting of indicators of pathology (for example, fatigue in cancer patients) with potentially treatable symptoms being overlooked.¹⁷⁰ Such observations have prompted speculation about the adaptive processes that may have served to confound such comparisons. The impact of response shift upon the validity of advance care directives is another area of interest.^{123 135}

Response shift as an intervention: Golembiewski noted that both recalibration and re-conceptualisation may be intervention goals in studies of organisational development.⁸⁹ Attempting to induce or 'teach' response shift may well be an important clinical intervention, for example, by encouraging increased discrepancy between objective and subjective indicators of well-being.^{10 161 169} Clinically, observed response shift may help patients by mitigating the effects of disease or therapy, for example, in cancer chemotherapy.¹¹⁸ Daltroy and colleagues assessed the discrepancy between self-reported and observed physical function in the elderly.¹⁴² Their findings suggested that re-calibration regarding functional ability was based on recent health difficulties. Providing individuals without such recent experience of declining function (e.g. through physical testing) could improve the agreement

between objective function and self-report. In this case, a response shift is being counteracted by the provision of a behavioural anchor. Daltroy and colleagues concluded that a more realistic appraisal of functioning may reassure patients and reduce demand for unnecessary healthcare. In contrast, Hagedoorn and colleagues suggested cognitive strategies based upon a positive relative evaluation may sustain quality of life in the presence of declining physical health.¹⁰⁷

Being cognisant of individual preferences and values may assist in designing therapeutic interventions which may engage response shifts in more acceptable and realistic areas of functioning. Sprangers and Schwartz illustrate this potential application with reference to a medical therapy for systemic lupus erythematosus with side-effects in one physical domain (gastro-intestinal) which may be easier for patients to accommodate to than if occurring in an alternate domain (facial disfigurement).¹⁰ Regimen adherence may be more likely if side-effects are located in areas that are less challenging to patient self-concept, and entail less response shift in more highly valued life domains.

Identifying changes in internal standards and conceptualisation may help patients with chronic disease integrate illness experience with changing life roles.⁹⁴ Schwartz described an intervention to improve coping skills in patients with multiple sclerosis which resulted in quality of life improvements despite reduced self-efficacy and neurological deterioration.¹⁰³ Schwartz attributed the divergence in physical and psychological well-being to response shift brought about by more positive coping strategies. Downward social comparison may have resulted from feedback from individuals with the same illness, but with poor experiences of coping. Lepore and Eton studied response shift in patients recovering from prostate cancer therapy and suggested that encouraging the amendment of life goals in the face of deteriorating urinary function may act to help restore a sense of personal control.¹⁰⁶

More globally, Norman and Parker suggested that the design of health-promotion interventions may benefit from awareness of effects at alpha, beta

and gamma levels of change.^{89 162} Thus, re-calibration and re-conceptualisation may need to be addressed if short-term effects of an intervention at an alpha level are to be maintained over a longer period of time. They also suggested that health re-conceptualisation resulting from an intervention designed to change one behaviour may generalise so that other health behaviours may also see change.

Adverse consequences of response shift: Several authors have provided a more cautionary note, arguing that inducing response shift as a therapeutic goal should only be considered after attempts at biological and physiological improvement have been exhausted.^{107 161} Another potential danger is relying upon adaptation (and subsequent response shift) as a justification for beneficial interventions with noxious side-effects. For example, Roos and colleagues note that pelvic exenteration is an extensive and mutilating procedure for patients with gynaecological and bladder cancer.¹²² That patient self-reported quality of life on several domains was no different from population values was seen as resulting from adaptation and a response shift. These authors took this as reassurance about pursuing the procedure. However, the observation that people adapt and cope after unpleasant experiences could be given too much weight in justifying invasive interventions. Thus response shift may be maladaptive (as well as adaptive) - for example, if re-calibration prevented otherwise appropriate use of health care services or therapies.¹⁰

Wilson describes potential applications in cases where the failure of normal and adaptive response shift has resulted in problems such as somatisation and hypochondriasis.¹⁶¹ Thus, therapies which aim to help patients think differently about bodily symptoms (e.g. to re-attribute meaning attached to sensations) may be effecting a response shift.^{118 161} Non-conventional therapies may benefit patients by helping them to think about their condition in different ways.

Medical decision-making: The relevance of response shift to medical decision-making has also been highlighted, for example, where preference

rating may vary by the health status of those completing the rating.^{95 166} This was the case work by Lenert and colleagues of Prospect Theory, where intermediate health states were valued nearly as much as good health by those respondents themselves in poor health.⁸⁷ In contrast, raters themselves in good health valued intermediate health states only slightly more than poor health states. The authors suggested that Prospect Theory and response shift are similar in that both provide mechanisms whereby utility values or preferences change as a function of changes in health. Health resource allocation based on rating by healthy people may discriminate against those who are ill. Preferences derived from proxy raters who do not experience adaptive processes may lead to different valuations from patients and change cost-effectiveness ratios.^{136 171}

4.5 Methods for assessing response shift

The 'thentest' design approach: From the literature review above, it is clear that the retrospective pretest-posttest (thentest) design is one of the earliest applied, and certainly more common, methodological approaches to directly assessing response shift. This approach involves prospective self-reports at two time-points (e.g. before and after an intervention) with an additional retrospective estimation of baseline level completed alongside the conventional 'posttest' assessment.⁹⁰ The difference between posttest score and retrospective rating of baseline is taken as measure of true change (i.e. unconfounded by response shift), since both assessments are made by the respondent using the same internal scale.

Bray and colleagues assessed the use of the thentest, and modelled different analytic approaches for measuring treatment effects.¹⁶⁵ They concluded that when response shift was present a thentest provides the most powerful method of analysis, as well as the best estimator of treatment effect. Given the prominence of this method, it is examined in more detail within chapter 9, where it is the approach selected for assessing re-calibration response shift in the current study. However, on its own it does not provide evidence regarding the mechanism producing response shift, only the size of

effect. As is also clear from the review, there are a variety of other methods, which are increasingly being applied to the study of response shift.

Assessing different forms of change: Millsap and Hartog conducted an early review of methods for distinguishing different types of change delineated by Golebiewski.¹¹³ Whilst a variety of techniques had been used to address re-conceptualisation, they used an operational definition of re-conceptualisation (Gamma change) as change in factorial structure from pretest to posttest, similar to Golembiewski's application of factor analysis.⁸⁹ In relation to re-calibration, they note how changes in 'ideal score' and retrospective pretest-thentest designs had been applied. Millsap and Hartog, themselves present the use of structural equation modelling as a method for addressing both re-calibration and re-conceptualisation.

Subsequently, to resolve the problem of multiple interpretations of change, Norman and Parker suggested an order of assessment aimed at ruling out firstly re-conceptualisation, then re-calibration and leaving the possibility of interpretation at the alpha level of change.¹⁶² The methods described represent broadly statistical and design approaches to assessment (table 4.5.1). The former involve mathematical manipulation of responses, the latter the use of new measures to determine the form of change occurring. Many of these approaches can be used to determine re-conceptualisation and re-calibration.

Table 4.5.1 Methods for detecting different forms of change

<i>Category</i>	<i>Method</i>	<i>Level of change assessed</i>	
		<i>Re-conceptualisation¹</i>	<i>Re-calibration¹</i>
Statistical	Transformation	✓	✓
	Coefficients of congruence	✓	
	Analysis of covariance structures	✓	✓
	Structural equation modelling	✓	✓
Design	Retrospective pretest-posttest	✓	✓
	Ideal scale		✓
	Criterion		✓

Adapted from Norman & Parker¹⁶²

1 Norman and Parker used the terms Gamma and Beta change (Re-conceptualisation and Re-calibration respectively)

Statistical approaches broadly compare factor structures of self-report outcomes at pre- and post-intervention, with lower congruence indicating re-conceptualisation. Whilst each approach can be applied retrospectively, reducing respondent burden, they require large samples. Furthermore, re-conceptualisation entailing the introduction of new domains not originally measured, would result in an incomplete assessment. Additional statistical approaches have been described, for example, the use of growth curve modelling, which offers the advantage of being able to model the timing and form (i.e. shape) of the change.¹¹¹ Multilevel modelling for assessing quality of life re-conceptualisation has also been described.¹⁵⁷ A key general consideration for statistical approaches is judging what level of re-conceptualisation would be clinically important. Design approaches share the similarity of an additional standard of measurement used to detect recalibration, for example the detection of changes over time compared to an 'idealised' state. The additional data collection required however, leads to the general criticism of design approaches that they add to respondent burden.¹¹³

162

Methods for quality of life outcomes: The methods reviewed by Norman and Parker only address the detection of re-calibration and re-conceptualisation, rather than exploring its nature. Therefore, they also advocated qualitative interviews (and, for example, content analysis). More recently, Schwartz and Sprangers reviewed a variety of methodological approaches for assessing response shift, specifically within the context of quality of life research. They distinguished between six broad approaches, for each of which they provide examples (table 4.5.2). Their expanded categorisation includes qualitative methods. Many of these approaches are not unique to, or even originally intended to address response shift but may be adapted to do so. Individualised methods, such as the Patient Generated Index and the Schedule for the Evaluation of Individual Quality of Life (SEIQoL) are included to explore re-prioritisation and re-calibration. The application of this method is described in chapter five.

Schwartz and Sprangers recognised that the distinctions made between classes of methods may be 'more apparent than real'. Furthermore, they also suggested that methods can alternatively be distinguished according to whether they involve self-referenced quality of life domain generation or standardised domains (i.e. ipsative or normative respectively). They also endorsed the sequencing whereby re-conceptualisation is addressed before re-calibration, but noted that partialing out the different forms of response shift may be more of an analytic convenience than a real reflection of events.

Table 4.5.2 Methods for assessing response shift in quality of life

Method	Example
Individualised	<i>Repertory Grid Technique, SEIQoL</i>
Preference-based	<i>Extended Q-TWIST, Preference mapping</i>
Successive comparison approaches	<i>Pairwise comparison, card sorting</i>
Design	<i>Thentest, ideal scale approach</i>
Statistical	<i>Covariance / factor analysis</i>
Qualitative	<i>Ideographic assessment of personal goals</i>

Adapted from Schwartz & Sprangers, 2000¹¹

Combining methods: Schwartz and Sprangers recommended triangulating approaches to address how re-calibration, re-prioritisation and re-

conceptualisation may combine to produce response shift.¹¹ A blend of qualitative and statistical approaches was used by Rapkin.¹³⁸ He described an idiographic approach to quality of life assessment in a study of people with AIDS which addressed re-calibration and re-conceptualisation. Following elicitation of personal goals using semi-structured paired interviews, variables identified via content analysis were submitted to regression modelling. Response shift was identified as significant interaction effects in an hierarchical regression analysis when observed changes in elicited goals moderate impact of events such as disease progression.

As suggested by Schwartz and Sprangers, a rigid classification and distinction between methods for assessing response shift is perhaps unrealistic.¹¹ For example, Donovan and colleagues examined the course of fatigue in women in receipt of radiotherapy for breast cancer.¹¹⁸ The predicted response shift was confirmed only for those women receiving prior chemotherapy. Those who had not received prior (toxic) treatment did not demonstrate a response shift and consequently reported increased radiotherapy-related fatigue. Thus, the study utilised an observational study design and also assessed statistically longitudinal trends.

An appraisal-based approach: With their revised theoretical model of response shift, Rapkin and Schwartz developed a semi-structured schedule for assessing the various appraisal parameters they had identified as contributing to quality of life self-report.⁹⁵ In part, it draws upon previous idiographic approaches in addressing personal goals to establish an individual's frame of reference.¹³⁸ Other elements address the three other appraisal parameters identified in the theoretical model, as well as a retrospective pretest-posttest. Furthermore, the schedule probes the respondent's view of discrepant answers from the thentest, and also gauges observed change in nominated goals.

4.6 Summary and implications for thesis studies

How important is response shift from a measurement or clinical perspective?

It is clear that response shift can substantially alter the size of a detected treatment effect (often underestimated) if it is not directly assessed.

Accordingly, treatment benefits may not be appropriately recognised. When baseline assessments exhibit ceiling and floor effects, response shift can be problematic. For example, baseline ratings of 'worst possible' may be recalibrated in retrospect to a less extreme value. Overall, longitudinal comparisons in the presence of un-measured response shift are difficult to interpret. Conversely, focusing upon response shift may reveal more about how individuals appraise constructs such as quality of life. Understanding the dynamism of quality of life over time may sharpen our ability to measure it in cross-section.

From a clinical perspective, it is clear that patients adapt to deteriorating health and that adaptation is associated with changes in internal standards, priorities and conceptualisation. Adaptation and response shift may even be expected clinically, and has driven attempts to implement response shift as an intervention. If the experimental group in a clinical trial reports an improvement, does it matter that this may reflect re-calibration rather than an inherent feature of the product (e.g. medication) being tested? However, it is more crucial in an explanatory trial where identifying the effective component of an intervention may be critical. Another potential difficulty is where patients adapt well to deteriorating health and continue to report satisfactory quality of life. If treatments of known efficacy are not considered for such patients, or such patients are accorded less priority in access to effective care this may serve to disadvantage them as they could still benefit from therapy.

Response shift has rarely been studied in patients with non life-threatening conditions, and my search failed to identify any studies reporting response shift in patients with musculo-skeletal or mobility problems (including mechanical knee problems). Qualitative (n=4) or individualised methods (n=1) have been used in only five of the identified empirical studies. The latter

article presented secondary analysis rather than data collected prospectively for this purpose.¹³⁵ In one of the qualitative studies, response shift was a reported outcome (observation) rather than the primary focus of the study.¹²¹ In two other qualitative studies (which used the thentest) respondents were asked to comment and expand upon discrepancies in prospectively and retrospectively derived answers about baseline state.^{144 145} Although qualitative and individualised approaches have been recommended for the assessment of response shift, only rarely have they been used prospectively for that primary purpose. The major component studies within this thesis will incorporate an individualised approach within a semi-structured qualitative interview. Allied to a quantitative thentest, the studies will evaluate the presence, nature and possible mechanisms of response shift in patients with mechanical knee problems.

Preface to chapter 5

The following chapter presents the overall design for the response shift work presented in the thesis. The subsequent chapters focus upon different aspects of that work but all utilise this common framework. Nevertheless, some of these subsequent chapters contain specific background, methods and discussion sections allowing them to be read largely autonomously.

Chapter 5: Research framework for evaluating response shift

***Abstract:** This chapter describes the methods framework used in theme two of this thesis for evaluating quality of life and response shift for a sample of patients with suspected or confirmed internal derangement of the knee. The epidemiology and nature of such knee problems is briefly described. The studies of the theme were conducted within the context of a randomised controlled trial evaluating the role of magnetic resonance imaging (MRI) for managing knee problems - the Direct Access to Magnetic resonance imaging: Assessment for Suspect Knees (DAMASK) Study. Therefore, the DAMASK study is briefly described, as is previous local work in this area. The research framework, centred upon individualised prospective assessment of quality of life across two interviews conducted six months apart is described. The detailed interview schedules; the assessments that comprise them; and the broad qualitative analytic strategy are described. Additional specific methods (where appropriate); and full results and discussion for each individual study are presented in chapters 6 to 12.*

5.1 Overview of study aims and presentation of response shift studies

The framework for the response shift studies is described below and addresses several research aims and objectives. These have been implemented, and are presented in subsequent thesis chapters as a series of component studies which are both qualitative and quantitative in nature. The broad qualitative approach to collecting and analysing interview data is described within this chapter. In addition, subsequent chapters will also describe background literature and methods where appropriate. The

framework interview study aimed to identify the presence of response shift in a sample of patients with ongoing knee problems who are at different stages of clinical management.

A preliminary stage of analysis aimed to describe and classify quality of life domains elicited within this sample of patients. This provided a baseline for understanding and exploring response shift. Subsequently, each component of response shift was assessed (re-calibration, re-prioritisation and re-conceptualisation). Re-calibration was assessed using a quantitative approach – the retrospective pretest-posttest design (or ‘thentest’). Re-prioritisation and re-conceptualisation were explored qualitatively.

Mechanisms and mediators of response shift were explored, including the impact of the clinical condition upon patient self-image. The relationship between respondent assessment of domain status (level) and the importance attached to that domain was specifically explored through both qualitative and quantitative means. The quantitative use of SEIQoL-DW cues and weights to assess response shift was reviewed and appraised using data from the current study. Finally, the impact of diagnostic and treatment delays upon patient well-being was explored qualitatively and the role of response shift discussed. Specific aims and objectives are summarised at the end of this chapter along with a key to which chapter they are addressed.

5.2 Background to study

Epidemiology and quality of life of patients with internal derangement of the knee

The overall response shift study will include patients with suspected or confirmed internal derangement of the knee. The morbidity survey for England and Wales (1995) reported by the Office of Population Censuses and Surveys

(OPCS) showed a recorded prevalence rate of 32 patients consulting per 10,000 person years at risk for internal derangement of the knee (table 5.2.1).¹⁷² This is similar to the combined rate for rheumatoid arthritis and other inflammatory polyarthropathies. Amongst the 16-24 year old age group the rate was higher (60 per 10,000), whilst in the 45-64 year age group the rate was lower (26 per 10,000). In addition, the consulting rate for sprains and strains of the knee and leg (ICD 844) was 80 per 10,000 person years, again with a peak rate in the 16-24 age group of 114.

Table 5.2.1 Annual consultation and new / first episode rates from OPCS morbidity survey, 1995 (rates per 10,000 person years at risk in England and Wales)¹⁷²

Age group	Patients consulting		New / first episode	
	Internal derangement (ICD9 717 ¹)	Knee and leg sprain / strain (ICD9 844 ²)	Internal derangement (ICD9 717)	Knee and leg sprain / strain (ICD9 844)
16-24	60	114	46	108
45-64	26	84	21	77
All	32	80	25	74

1 Includes derangement of lateral and medial meniscus

2 Includes lateral and medial collateral ligament, cruciate ligament

Generic outcomes: Hollingworth and colleagues assessed quality of life in 332 patients before and after MRI referral for knee problems at various anatomical sites (including 46% meniscal, 18% patellofemoral joint, 12 cruciate ligament).¹⁷³ Both men and women exhibited significantly poorer health status compared to the general population for physical functioning; role limitation (physical); pain; and social functioning (table 5.2.2). To a lesser extent general health; energy and vitality; and mental health were also affected. Significant differences between discharged and not discharged patients were found on three SF-36 sub-scales (physical functioning; role limitations and pain) six months later. However, significant impairment on five sub-scales was still evident for patients, whether discharged from secondary care or not.

In the same study, baseline impairment on three dimensions of EQ-5D was evident for patients compared to the general population (pain and discomfort; usual activities and mobility). Despite improvement after six months, patients still reported impaired scores on the same sub-scales.

Table 5.2.2 Impaired health status / HRQoL associated with knee problems compared to general population ¹⁷³

SF-36 sub-scale	Impairment compared to general population		Improvement at six months
	Baseline	Six months	
<i>SF-36 domains</i>			
Physical functioning	Yes	Yes	Yes ²
Role limitations (physical)	Yes	Yes	Yes ²
Pain	Yes	Yes	Yes ²
General health	Yes ¹	Yes	
Energy and vitality	Yes ¹		
Social functioning	Yes	Yes	Yes
Role limitations (emotional)			
Mental health	Yes ¹		Yes

1 Difference found only with larger sample combining males and females

2 Significant difference between discharged and not discharged patients

Specific outcomes: Generic health status measures have, therefore, been successful at detecting sub-optimal well-being. Furthermore, there are a large number of specific scales which have been applied to a variety of knee-related problems (including osteoarthritis, patellofemoral pain and injuries to the meniscus, ACL and cartilage).¹⁷⁴ Borsa and colleagues distinguished between two broad approaches to the assessment of disability: performance-based and patient-reported methods.¹⁷⁵ They found the latter proved better predictors of patients' subjective rating of disability in ACL deficient individuals. These measures focused upon symptoms and functional limitations related to activities of daily living.

Amongst the specific measures produced, Lysholm, described the development of a scale for assessing patients with knee ligament injuries which incorporated the notion of knee instability.¹⁷⁶ A second example of a

specific scale was provided by Mohtadi who described the development of a measure of the chronic problems associated with ACL deficiency.¹⁷⁷ The resultant ACL-QOL comprised five domains: symptoms and physical complaints; work-related concerns; recreation and sport concerns; lifestyle concerns; social and emotional concerns.

Rockborn and colleagues found that, in routine practice, patients undergoing uncomplicated arthroscopic meniscectomy may frequently be able to return to work within a week.¹⁷⁸ Up to two-thirds of patients reporting complete recovery, did so within four months although for some, recovery time was longer.¹⁷⁸ Wexler also assessed health status for patients followed up between 24 and 98 months (mean = 55) after anterior cruciate ligament reconstruction.¹⁷⁹ Compared to general population levels, patients reported worse mental health scores but better role physical and general health scores (SF-36).

Context to response shift study: The DAMASK Study

The response shift studies conducted within this thesis utilised the general framework of the DAMASK study, which itself built upon local work assessing primary care access to specialist imaging technology for musculo-skeletal problems. Direct access to MRI for patients with suspected meniscal or ligamentous tears has been available to general practitioners in Cardiff since January 1993.¹⁸⁰ The mean waiting time for the investigation was 19 days for primary care patients when the service was initiated. Access has been guided by the development and implementation of clinical referral guidelines which have been shown to improve clinician's knowledge and the proportion of appropriate referrals.¹⁸¹⁻¹⁸³

Funded by the UK Medical Research Council, the on-going DAMASK study is assessing the role of MRI in UK primary care for the management of patients with suspected internal derangement of the knee. Specifically, the study is evaluating how the technology influences diagnosis and management of knee patients; how it affects patient outcomes; and whether it reduces costs to the individual, the health service and society.

There are two principal components to the DAMASK Study: a randomised controlled trial (RCT) and an observational survey. At the time of writing the DAMASK trial is currently being conducted in the north of England, Scotland and Wales. Patients in the trial are recruited in general practice and randomly allocated to receive either MRI investigation followed by an orthopaedic review or direct referral to orthopaedics. Patients are eligible for trial inclusion if they are aged between 18 and 55; are suspected of suffering from internal derangement of the knee (meniscal or ligament injury); and their general practitioner is considering orthopaedic referral. Outcome assessment occurs at baseline and at six, twelve and twenty four months after the patients is recruited to the trial. Outcomes include general measures of health status and health utility (SF-36 and EQ-5D) as well as a specific outcome measure.¹⁸⁴ The latter instrument was developed by the trial team, including myself. This involved semi-structured qualitative interviews and focus groups with patients recruited in Cardiff and York. The focus of instrument development work was an assessment of how patient quality of life was affected by their knee problem.

The DAMASK survey was conducted only in Cardiff, South Wales and included similar patients, but aged up to 65 years. Cardiff is a clinical centre with established access to MRI for general practitioners. In the survey, newly referred patients were recruited from the departments of radiology (University Hospital of Wales NHS Trust) and orthopaedics (Llandough Hospital), Cardiff and Vale NHS Trust. These patients were followed up by postal questionnaire using the same time-points as the RCT. In addition, patients about to undergo arthroscopy at Llandough Hospital were also surveyed. The same outcome measures are being used in the trial and survey questionnaires.

In the DAMASK study, to conduct the interviews for outcome measure development and to conduct the observational survey, access to patient details was negotiated with clinical collaborators from both Radiology and Orthopaedics. The Response Shift interviews of this thesis were conducted immediately prior to the DAMASK postal survey (and some time after the

outcome measure development interviews). The patient identification system was developed for both the DAMASK study and the Response Shift studies (although sampling for the latter was different). Details of the system and more generally the design of the Response Shift study are now described below.

5.3 Response shift studies: design

Overview: The sample included patients referred by their general practitioner to the Department of Orthopaedics, Llandough Hospital and Department of Radiology, University Hospital of Wales, Cardiff, with suspected internal derangement of the knee. A purposive sample was selected to reflect different ages; stages in management process; and gender – thus, forming eight study groups. The aim was to interview five patients per group, a total of 40 patients. Patients met the general entry criteria for the DAMASK survey in Cardiff (aged between 18 and 65) but were recruited prior to the start of the main survey sample. Patients approached and agreeing to study inclusion were interviewed using a semi-structured schedule, incorporating an individualised quality of life assessment (SEIQoL-DW). The same patients were followed up by interview six months later using a modified version of the original interview schedule. Data collected during the two interviews were both qualitative and quantitative in nature and allowed exploration of each of the three theoretical components of response shift (re-calibration, re-prioritisation and re-conceptualisation).

Ascertainment of patient sample: Patients were identified from the departments of Radiology and Orthopaedics and included recent referrals for a knee problem (waiting list group) and those about to undergo a therapeutic arthroscopy (arthroscopy group). In Radiology, all new requests for MRI of the knee were photocopied by the MRI coordinator and passed to MR. Patients matching the general entry criteria for the study were selected for approach. In Orthopaedics all new referrals to each of the three collaborating orthopaedic surgeons were photocopied and set aside for the research team.

The surgeons assessed the referral against the clinical eligibility criteria for the study. The researcher (MR) confirmed the general eligibility of these patients (for example, within the indicated age range) prior to an approach being made.

The process for identifying patients about to undergo arthroscopy was different. Patients were identified from the clinical diaries for each collaborating surgeon. Patients could have been entered into the diary up to a month in advance of their planned operation. A member of the DAMASK research team visited weekly and in some cases the intervals between appointment being made, operation date and visit date may have been quite short. The opportunity to approach and interview such patients was therefore limited. Patients identified only after their operation date were not approached.

Further details for identified arthroscopy listed patients were provided by the Department of Orthopaedics to determine the nature of the operation and to ensure patients met the eligibility criteria for study entry. This included, where available, relevant clinical read codes (table 5.3.1). Details for all patients (arthroscopy and waiting list groups) were entered onto a study administration database (written in Microsoft Access by MR).

Table 5.3.1 Read codes used to select arthroscopy listed patients

W72 (Prosthetic Replacement of Ligament)
W73 (Prosthetic Reinforcement of Ligament)
W74 (Other Reconstruction of Ligament)
W75 (Open Repair of Ligament)
W76 (Other Operations on Ligament)
W77 (Stabilising Operations on Joint)
W81 (Other Operations on Joint)
W82 (Therapeutic Endoscopic Operations on Semilunar Cartilage)
W83 (Therapeutic Endoscopic Operations on Articular Cartilage)
W84 (Therapeutic Endoscopic Operations on Other Joint Structure)
W85 (Therapeutic Endoscopic Operations on Cavity of Knee Joint)
W86 (Therapeutic Endoscopic Examination of Knee Joint)
W90 (Puncture of Joint)

Purposive sampling: The sample was formed on the basis of gender (male: female); age (<40: 40+); and stage of clinical management (arthroscopy: waiting list) – a form of stratified purposeful sampling.¹⁸⁵ Sampling by gender was intended to maximise variation in the sample and allow exploration of differences by gender. The same was true for age. It is possible, for example, that for older patients, internal derangement of the knee may be viewed as a condition associated with ageing and its impact upon lifestyle may be viewed as less important and more easily accommodated. For younger people, the impact of the knee problem upon lifestyle and well-being may be of a different nature and less acceptable. Patients with internal derangement of the knee are unlikely to improve physically without intervention. Therefore, purposive sampling by stage of management allowed exploration of whether any changes in quality of life (for example, re-conceptualisation) were evident in a period of steady state (recently referred patients on a waiting list) or clinical intervention (arthroscopy). Although patients receiving arthroscopy were likely to experience physical improvement following intervention, this would not necessarily be the case for all. Nevertheless, they were more likely to experience some physical change (either positive or negative) compared to the waiting list group.

In consultation with an experienced medical sociologist (Professor Roisin Pill), the size of the study sample was derived on the basis of the three criteria specified above, and likely coverage of the phenomena under study.¹⁸⁵ A total of five patients per group provided a sample of forty patients (table 5.3.2) and eighty interviews in all. It also allowed 20 patients (and forty interviews) per study sub-group (for example, males and females). On the basis of our previous qualitative work, twenty interviews were considered to be sufficient to allow category saturation. Thus, the sample size chosen maximised sample variation, enabled group comparisons (where required) and included some redundancy to account for sample non-response and failure to follow-up.

Table 5.3.2 Response shift interviews: purposive sample groups

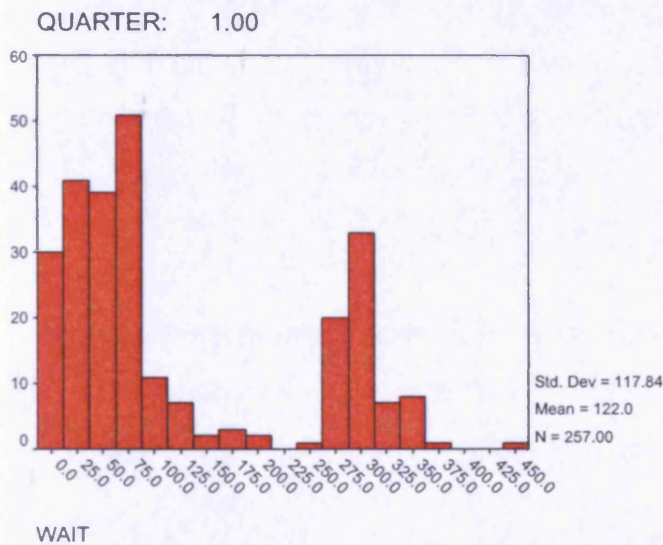
		<i>Gender</i>				
		Male	Female	Male	Female	
Age	< 40	n=5	n=5	n=5	n=5	
	> 40	n=5	n=5	n=5	n=5	
		<i>Waiting list</i>		<i>Arthroscopy</i>		<i>Stage of management</i>

Patient approach: All patients meeting entry criteria were sent a letter of approach, information sheet, consent form and pre-paid envelope. The approach letter was presented on the headed notepaper of the host clinical department (Orthopaedics or Radiology) and signed by the relevant clinical collaborator. Patients returning signed consent forms were telephoned by MR to arrange an appointment. New patients were approached until the required sample size was achieved. Patient approach letter, information sheet and consent form are shown in appendices IV, V and VI.

Timing and conduct of interview: Patients were interviewed on two occasions, six months apart. The first interview for the waiting list group (recently referred patients) was soon after the original referral. Contemporary routine waiting times for orthopaedic out-patient appointment was over two years. Similarly the current waiting time for MRI of the knee for patients referred from general practice ranged from six to twelve months (figure 5.3.1). Therefore, these patients were unlikely to have had any secondary care intervention (either diagnostic or therapeutic) by the time of their second

interview. The first interview for arthroscopy patients was conducted in the month prior to their arthroscopy. Their follow-up interview was six months after this and, therefore, approximately five months post-surgical. All interviews were conducted by MR.

Figure 5.3.1 Interval in days between general practice referral and MR scan of the knee (first quarter 2001)



Data management: All interviews were audio-recorded with the permission of the patient using a mini-disc recorder. Notes were taken during the course of the interview and summary notes subsequently recorded on the study database. Interviews were transcribed for subsequent analysis. Qualitative analysis was supported by the QSR NUD*IST software programme. Quantitative data management and statistical analysis was conducted using SPSS for Windows software (Version 11.0.1) and Minitab (Version 13.32).

Ethical approval: The interview study was approved as part of the main DAMASK study by Northern and Yorkshire Multi-Centre Research Ethics Committee

5.4 Principal outcome measures

Individualised quality of life measures: Schwartz and Sprangers described several individualised instruments that may be used to assess response shift in quality of life, including the Schedule for the Evaluation of Individual Quality of Life (SEIQoL) and the Patient Generated Index (PGI).^{11 186} Both instruments were designed to assess quality of life from the perspective of the individual respondent, and each approach involves three stages.¹⁸⁷ Firstly, important areas of life are nominated by the respondent. In the second stage, each nominated area is scored according to its current level. The final stage requires the respondent to assign a weight to each area according to its relative importance.

The original version of SEIQoL uses judgement analysis to derive weights for quality of life domains (termed cues).¹⁸⁸ The process involves an individual using a visual analogue scale to rate their perceived quality of life under 30 different randomly generated hypothetical scenarios, each of which incorporates the respondents own quality of life cues.¹⁸⁹ This time-consuming process presents the respondent with a high level of burden and results in some non-response.^{189 190} A shorter version of the assessment (SEIQoL-DW) uses a direct weighting procedure in the form of a 'dynamic pie chart'.^{189 191 192} Waldron and colleagues reported completion times for SEIQoL-DW in a sample of patients with advanced cancer which were on average 15 minutes compared to 40 for the approach using judgement analysis.¹⁹³ In a sample of healthy volunteers Browne and colleagues reported completion times of less than five minutes and between 10 and 30 minutes for the direct weighting and judgement analysis approaches respectively.¹⁸⁹ Completion time required for SEIQoL-DW compares well with other individualised approaches such as the Repertory Grid technique (e.g. up to sixty minutes); and other standardised measures (for example, Sickness Impact Profile).^{194 195} It was, therefore, the short form of the SEIQoL tool that was considered for use in this study.

Joyce and colleagues described the use of SEIQoL (by Dec 2000) in over a dozen countries and in more than 200 studies.¹⁹⁶ They concluded that SEIQoL was suitable for use in any clinical population and that it is one of ten quality of life methods listed by WHO (and the only one for individualised quality of life). Neudert and colleagues have also reported that SEIQoL-DW causes less distress to patients upon completion compared to standardised health status measures such as the Sickness Impact Profile.¹⁹⁵

Choice of SEIQoL-DW over PGI: As stated above SEIQoL and PGI share common features. Both are generic (in that they may be applied across different patient groups, diseases and cultures); they both allow the patient to define areas of importance; and both use prompt lists if cues can not be nominated.¹⁹⁷ It has been argued by MacDuff, that the very dynamism of quality of life may potentially hinder reliability assessment of individualised measures, although he acknowledged that adequate test-retest reliability has been demonstrated for some measures.¹⁹⁸ Furthermore, a more recent review of individualised approaches by Patel and colleagues concluded that both SEIQoL and PGI had adequate test-retest reliability.¹⁹⁴ They also reported that SEIQoL has good internal reliability and unlike PGI, SEIQoL has high content validity.

SEIQoL was designed for application via semi-structured interview whilst the PGI was initially intended to be self-administered, thus enabling use in postal questionnaires.^{186 192} As the chosen outcome measure was to be completed during the course of a semi-structured interview, SEIQoL-DW was considered more appropriate for this purpose. Completion problems have also been reported with PGI and shown to be associated with certain socio-demographic characteristics (e.g. respondent education level and form of household tenure) leading to response bias.^{186 199} Lindblad and colleagues considered the PGI to be cognitively complex, and as such was a disincentive to selection in the current study.¹⁹⁷

The scope of the two measures also differs. The Patient Generated Index aims to directly assess the impact of a specified medical condition upon an individual's quality of life. Hence, in its application the condition of interest is provided and used as the anchor for deriving content areas (box 5.4.1).¹⁸⁶ A sixth area or activity may be used to encompass issues not already mentioned, including those unaffected by the specified medical condition. In contrast, for SEIQoL-DW the focus is broader and does not ask the respondent to reflect on any particular concern.¹⁹² This has allowed it to be used to collect data on non-patient groups for comparative and normative purposes.^{191 200} Even within patient groups, 'health' may not be nominated as one of the five quality of life area cues.^{191 201} As accommodation to on-going knee problems may result in a change in focus away from areas of life affected by the knee, it was decided that a tool not directly focusing upon the knee would be more appropriate. Finally, there is also evidence that PGI scores more closely relate to physical functioning, whereas SEIQoL-DW scores relate more to measures of general health and vitality.²⁰²

Box 5.4.1 Extract from Patient Generated Index¹⁸⁶

At this stage, we would like you to think of the different areas in your life, or activities in your life that have been affected by your *name of condition* in the last month.

Completing and scoring SEIQoL-DW: SEIQoL-DW is administered by an interviewer trained using the instrument manual.^{191 192} Completion proceeds through three stages. Firstly, the respondent is asked to consider areas of life that are important to their overall quality of life. They are asked to nominate five such cues. Most respondents are able to nominate five cues, but if not, a standard list of prompts is provided.^{192 201} However, the instrument may still be completed if less than five cues are nominated.¹⁹¹

When eliciting the cues, the interviewer is required to record both the label used to describe the content area as well as a description of it.¹⁹² This is particularly important when the respondent is to be re-assessed subsequently

to determine if the same cues are being described. In the SEIQoL-DW administration manual, O'Boyle and colleagues state that clarifying the description of the cue also facilitates the summarising of cues when aggregating data from a group of respondents.¹⁹² The label recorded is that provided by the respondent rather than the interviewer.

The second stage requires the respondent to rate each elicited cue on a vertical scale anchored at either end by descriptors of 'worst possible' and 'best possible'. The scale is 100mm in length and allows the respondent to apply their own criteria for reporting their status. The third stage asks respondents to quantify the relative importance of each cue using a stacked set of five laminated discs. The discs are centrally mounted and each is labelled with an elicited cue label. The respondent is asked to rotate the discs so that the proportion of each disc showing represents the relative importance of the cue. The relative proportions are measured using a scale on the external edge of the disc.

Although intended for measuring change in an individual, an 'index' quality of life score may be obtained for the purpose of group comparison.¹⁹² This is calculated by multiplying each cue level (from stage two) with the corresponding weight (from stage three) and aggregating the resulting values. SEIQoL-DW has also been applied in modified form, for example by Wettergren and colleagues.²⁰³ They allowed respondents to nominate as many cues as they wished; rated current cue levels on a seven point scale; and produced overall scores which could range from one to seven. Other variations in completion include an approach by Tovbin and colleagues in which patients ranked cues rather than precisely determining weights.²⁰⁴

5.5 Assessing response shift using SEIQoL-DW

When administering SEIQoL in a longitudinal study with repeat assessment, the instrument's administration manual recommends that cues are elicited de novo at each assessment point (which, therefore, allows for different cues to

be elicited).¹⁹² If the two sets of cues are different, those originally elicited at baseline are provided for the respondent to rate and the full SEIQoL procedure is repeated to enable direct comparison across time-points.¹⁹² Such a procedure was described by O'Boyle in a prospective study of patients undergoing hip replacement.²⁰¹

Instability of cue nomination was used as evidence of re-conceptualisation and the primary indicator of response shift in the present study.¹³⁵ Therefore, SEIQoL-DW was implemented at follow-up using the cues elicited at follow-up, regardless of baseline cues. This approach to response shift assessment was described using previously collected data by O'Boyle and colleagues.¹³⁵ Similarly, in a prospective study of the healthy elderly Browne and colleagues also obtained levels and weights for cues derived at follow-up, regardless of their presence at baseline.¹⁹⁰ This should enhance the validity of the second assessment by using cues considered relevant by the respondent at that point in time. This may be especially important in those for whom adaptation or accommodation has resulted in a response shift. However, cues nominated only at baseline were still addressed in a subsequent section of the interview using probes to determine the nature of change that may have occurred. This formed part of the individual level qualitative analysis.

O'Boyle and colleagues suggested that, for cues nominated both at baseline and at follow-up, a change in weight was indicative of re-prioritisation response shift.¹³⁵ Nevertheless, these authors noted potential difficulties with this approach. The derivation of weights via the judgement analysis of the original SEIQoL method may be confounded by scale re-calibration – a change in weight may not simply be the result of change in values.¹³⁵ The direct assessment of weights in the shorter form of SEIQoL reduces this problem by not using the global rating which is the basis of the judgement analysis. Furthermore, in the current study the individual level analysis sought to identify supportive evidence of re-prioritisation, in addition to simple observation of consistency of weights. A second problem noted by O'Boyle and colleagues was that assigning weights to one cue was not independent of the weighting of remaining cues as total weights are constrained to unity.

Whilst O'Boyle and colleagues have proposed the use of free weights as a more appropriate means of assessing such change, this would modify significantly the completion of SEIQoL.¹³⁵ Therefore, in the present study greater emphasis was put on changes in ranking associated with derived weights, and on probing respondent explanation or understanding of such differences.

5.6 Interview schedules

The baseline and follow-up interviews followed semi-structured schedules (appendices VII and VIII). Although their content varied, both were built around the SEIQoL-DW interview schedule and shared a common structure (table 5.6.1). The design of the interview guides was informed by my previous experience of interviewing and running focus groups during outcome measure development for the DAMASK trial.

Baseline interview

The baseline interview commenced with a general introduction from the interviewer explaining his affiliation and role; the aim of the study; how the patient was identified; and why they had been approached for study participation. The patient was informed about the content of the interview and its likely duration. Following a check that the patient was content with the planned process, the interview continued with administration of the SEIQoL-DW. Additional probes were used at each stage of its administration to expand upon the cue definition; reasons for observed cue level; and cue ranking. Descriptive data from cue elicitation and at subsequent points of the interview were used to determine whether apparent differences were credible as representing response shift change.

Patient narrative: Following SEIQoL-DW administration, patients were asked to reflect upon how the content, or their weighting of cues, may have changed during the preceding year. The remaining 'half' of the interview involved a more narrative, and less structured discussion of the patient's knee problem

and how it had affected them. Several set probes were used and included asking patients to describe the onset of the knee problem and what they felt caused it. They described the nature of the problem both in terms of symptoms and also its functional impact. The consequences for their quality of life, including the already elicited SEIQoL cues, were probed. This was similar to the approach applied by Wettergren and colleagues who used a disease-specific module with a modified version of SEIQoL-DW to determine the impact of disease (Hodgkin's lymphoma) on patient's lives.²⁰³ Another probe in this section asked patients to reflect upon how they thought they would be affected by their knee problem in the future.

A key question in this section asked patients to consider whether the knee problem, and its impact upon them, had changed the way that they felt about themselves. This question was intended to explore whether the knee problem and the patient's response to it had affected their self-image. Patients were asked to complete the EQ-5D health utility scale which provides a standardised measure of health status. The interview finished with the interviewer summarising the main observations and checking whether the patient had anything else to add.

Table 5.6.1 Overview of baseline and follow-up interview schedules; variations in methods between interviews; and indication of relative contribution to study analyses

Schedule section	Description of section content	Interview variation: whether included (Yes / No) and differences in method		Planned analysis (see table 5.8.1)
		Baseline	Follow-up	
Introduction	Interviewer introduces self; explains reason for study; how and why patients approached; process and content of interview. Checks patient happy with process	Yes	Yes	
SEIQoL-DW	Expanded SEIQoL interview administration – elicitation of content cues, current cue level; and weight using DW procedure. Additional cue prompts provided if required. Standard probes used: <ul style="list-style-type: none"> • Expanded cue definition • What is affecting rating of cue level • Knee related causes (if apparent) • Explanation for ranking and ease of process 	Yes	Yes	a, b, c, d, e, f, h, i
Direct assessment of content change	Prompt to explore changes in cue content and weighting over the previous year	Yes	No	a, b, c, d, e
Health transition	Knee-specific health transition items (physical, mental, general) since the baseline assessment (6/12)	No	Yes	g

Table 5.6.1 (cont)

Schedule section	Description of section content	Interview variation: whether included (Yes / No) and differences in method		Planned analysis (see table 5.8.1)
		Baseline	Follow-up	
Knee narrative	Prompted patient narrative about aspects of their knee problem with emphasis varying at baseline and follow-up, but both addressing impact of problem, consideration of causation; and the future.	Yes – including onset; nature of problem	Yes – including coping; understanding of problem	e, j
Self-image	Direct assessment of impact of knee problem upon self-image	Yes	Yes	e
Direct response shift probe	Direct assessment of observed change in cue content or weighting	No	Yes	a, b, c, d, e
LOT-R	Standardised assessment of dispositional optimism	No	Yes	g
EQ-5D	Standardised health status assessment	Yes - pretest	Yes - posttest followed by retrospective pretest (thentest)	f, g
Close	Summarise, probe for any further issues, close	Yes	Yes	

Follow-up interview

The follow-up interview schedule followed the same broad structure as the baseline interview. Patients were reminded of the aims of the study; and the process and content of the interview. The SEIQoL-DW was administered including the additional probes that had featured in the baseline interview. The patient was then asked to complete three specific health transition items. These asked the respondent to describe how their knee was affecting their daily activities; how much time they spent worrying about their knee; and how their knee was in general compared to six months ago. The first two items provided a seven-point adjectival response scale and the third item a 15-point adjectival response scale. The three items used had been developed for use in the DAMASK study. In the interview they mainly served to orientate the respondent to reflect on how things may have changed for them since the baseline interview. However, they were also used in the analysis of recalibration response shift.

Patient narrative: The next section of the interview schedule was a more narrative description of the impact of the knee problem upon the patient since the baseline interview. Respondents were asked to describe how they were currently affected and what had occurred since the first interview. Specific probes included how they had coped with, or adjusted to, the knee problem; and its impact upon their life. They were asked to reflect on whether their understanding of the cause and nature of the problem had changed since they were first interviewed. As in the baseline interview, they were asked to describe how they expected to be affected by their knee problem in the future. They were asked to consider how their knee problem, and its impact upon them, may have changed the way that they thought about themselves and their approach to life.

Direct probe for response shift: Patients were then asked directly about cue content or weighting changes between assessments (table 5.6.2). The interviewer had details of the elicited cues and weights at baseline. The probing process was structured in a hierarchy, whereby the probe the

interviewer used was dependent upon whether there was an observed change in SEIQoL cue content; cue weighting; or no change. If the patient had mentioned different cues at follow-up they were asked to describe why they thought their nominated cues had changed. This was an opportunity to determine the credibility and validity of observed differences between assessments (e.g. that differences were not simply due to memory problems). It also provided an opportunity to probe for the patient's understanding about why things may have changed. If cue content was consistent across assessments, the patient was asked about any observed changes in weighting or relative ranking of elicited cues. Again this was an opportunity to explore reasons for such changes. Finally, if both cue content and weighting were consistent across ratings, patients were asked if they thought that any cues mentioned at baseline had changed and why. In this case patients were not informed in advance of the stability of their cues.

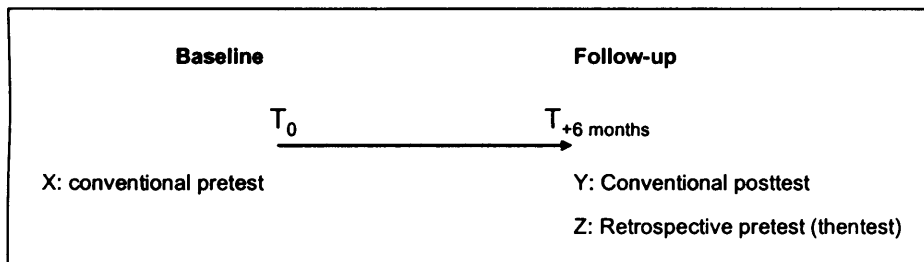
Table 5.6.2 Process hierarchy for direct probing of response shift

	Observed change (baseline: follow-up comparison)	Patient informed of baseline cues	Interviewer probe	Response shift component explored
i)	Cue (content / label)	Yes	Explore reasons for change	Re-conceptualisation
	<i>If no cue content / label change interviewer selects ii)</i>			
ii)	Cue weight	Yes	Explore reasons for change	Re-prioritisation
	<i>If no cue weight change interviewer selects iii)</i>			
iii)	None	No	Explore patient perception of change	Re-conceptualisation / re-prioritisation

Assessing dispositional optimism, and re-calibration: To assess the potential role of optimism as an antecedent in response shift a measure of dispositional optimism was included.²⁰⁵ The revised Life Orientation Test (LOT-R) comprises ten items, although only six are used for compiling a scale score. These likert-like items have five response steps ranging from strongly disagree to strongly agree and form a scale score which may range from zero to 24. It was decided to include the LOT-R in the follow-up interview after some initial baseline interviews during which some respondents referred to

optimism. The LOT-R was read out by the interviewer. The patient responded verbally with the aid of a pre-printed card containing response options. The final assessment was the health utility measure, EQ-5D, which was applied as part of a retrospective thentest (figure 5.6.1 and described more fully in chapter 9). This was given to the patient to complete for how they felt currently (referred to as a conventional posttest – denoted by Y). Once completed, they were then asked to complete it again for how they felt six months previously at the baseline interview (referred to as a ‘thentest’ and denoted by Z). The interview concluded with the interviewer summarising what had been described in the interview and checking whether the patient felt there was anything else to add.

Figure 5.6.1 Assessment time-points for retrospective pretest-posttest (thentest) evaluation of recalibration response shift (using EQ-5D)



Piloting the interview schedule: The draft baseline schedule was piloted upon two patients who had previously been interviewed by MR as part of the development process for the specific outcome measure used in the main DAMASK study. The interviews were conducted as intended for the main interviews and were recorded and transcribed. Data from the pilot were reviewed by MR and two other researchers, Roisin Pill (medical sociologist) and Chris Butler (general practitioner). The baseline schedule was finalised following piloting. The follow-up schedule was finalised following completion of the baseline interviews.

5.7 Qualitative analytic approach

Analysis of interview data was planned at both an individual level and group level, represented by broadly qualitative and quantitative strategies respectively. Following Schwartz and Sprangers, a change in the content of the elicited cues between the baseline and follow-up assessment for each patient was taken as an indication of re-conceptualisation.¹¹ Similarly, a change in the weight assigned to cues between assessments was taken as an indication of a re-prioritisation. The analysis was initially descriptive moving towards conceptual ordering. One intended outcome from this process was the generation of hypotheses about the nature of relationships between observed factors that emerge from the data. The general analytic strategies employed in the qualitative analysis are described below.

Using memos: Maxwell described three broad options for qualitative analysis: (i) creating and using memos, (ii) categorising strategies and (iii) contextualising strategies. He further described how qualitative analysis should ideally combine each of these strategies.²⁰⁶ Therefore, the individual level qualitative analysis firstly involved the use of memos to record initial reflections upon the interview. Notes were made immediately after each interview to record observations about the interview context; completion of the study materials; and initial thoughts about interview content. These data were recorded upon the study administration database. Using memos to start making sense of emerging data represents both data collection and also initial stages of the analysis process.¹⁸⁵ Memo creation also featured in the computer-assisted analysis of the data. Transcribed interview data were imported into a qualitative software package (QSR N6) for analysis.²⁰⁷ Memos were used for describing derived nodes and emerging categories.

Categorising strategies: The principle categorising strategy employed was the identification of themes from the data and the development of a descriptive framework. The structure of the overall schedule was partly driven by the requirements of SEIQoL-DW and, to a lesser extent, the other

quantitative assessments being conducted. However, even within this administration of SEIQoL-DW, the aim was to allow greater exploration of cue content and factors associated with their provision, rating and weighting. The resultant semi-structured interview guide enabled an analytic framework approach to data organisation whereby responses to specific topics in the interview were grouped together.¹⁸⁵ The principle a priori grouping areas for this study were as follows:

- Descriptions of quality of life area *content* (stage one of SEIQoL-DW) – used to develop a framework of quality of life areas (chapter 7)
- Descriptions of change in quality of life content areas (re-conceptualisation) - explored to identify and classify emergent themes (chapter 8)
- Descriptions of change in *ranking* (re-prioritisation) in quality of life areas (chapter 8)
- Assessment of treatment delay upon quality of life (chapter 12)

Whilst the interview schedule provided specific questions or subsequent prompts to ascertain the above data, relevant data also emerged from other parts of the interview. This included data observed directly, for example, via spontaneous comments and reflections by patients about change in the importance of a quality of life area. It was also indirectly observed by analytic comparison by the researcher of quality of life cue content (and weighting) between the baseline and follow-up interview. These distinctions are noted in the results where appropriate.

Both baseline and follow-up interviews represented a pre-structured approach to qualitative data collection and analysis.²⁰⁶ This approach is particularly useful where the research questions are well delineated.²⁰⁸ Maxwell usefully distinguished between the degree of structure used in data collection and how the structure is used in the analysis.²⁰⁶ He also described how within a single method, flexibility may be retained to allow for more in-depth questioning and analysis in response to emerging insights. This flexibility is built into the

'second half' of the research interviews in particular and permitted a more inductive analysis of the impact of the knee problem and its management upon the respondents' well-being. That the content of the respondent's comments reflected the structure of the schedule was recognised and assessed accordingly in the analysis.

The semi-structured interview guide enabled comparison between individuals and between groups being studied (for example, male and females).²⁰⁶ The degree of pre-structuring, and the way that it was employed in the analysis (a form of 'pre-analysis'), simplified the analytic work.

Content analysis: As noted earlier, Schwartz and Sprangers suggested the use of content analysis within both individualised and qualitative approaches to response shift.¹¹ There are a number of approaches to content analysis. Silverman described a quantitative approach whereby categories are established and then the occurrences of these within a text are counted.²⁰⁹ Such coding schemes may inhibit analysis if they are applied too rigidly, to the detriment of activities that are harder to fit within the schema.²⁰⁹ Silverman was concerned about the rapid derivation of categories which ignore the way that respondents themselves choose and use categories to frame their activities.

In contrast, Patton described content analysis as qualitative data reduction to make sense of the data.¹⁸⁵ This involves developing manageable coding systems to categorise and label primary patterns in the data. As the aim of the analysis was primarily to explore the nature of patient experience, rather than quantify it, Patton's analytic approach was applied to the study interviews. In data collection, probing of interviewees regarding their responses represented an attempt to understand from their perspective their conceptualisation of cue content, level and weighting.

Coding strategy: Initial read through of post-interview memos and transcribed interviews initiated the development of inherent themes.¹⁸⁵ Category development utilised analytic convergence – identifying what data

elements fitted together through recurring regularities. The degree to which data elements within a category formed a cohesive whole (internal homogeneity) and the degree of discrimination between categories (external homogeneity) was assessed. Wettergren and colleagues used a modified version of SEIQoL-DW to assess the impact of Hodgkin's lymphoma, and its treatment, upon quality of life.²⁰³ They used a pre-existing coding framework to initially structure the elicited cues. However, in the current study, although the overall interview was semi-structured, analysis of patient descriptions of quality of life was inductive in that a pre-existing framework was not applied. Therefore, patterns and themes were allowed to emerge from the data

Contextualising strategies: Contextualising analysis aims to describe the relationships between various elements under investigation and present these within a coherent whole.²⁰⁶ One approach to contextualising is the use of case studies. The unit of analysis in such a case study may be a critical incident (such as having a knee operation) or an individual patient.¹⁸⁵ The latter was used in the present study. Summary case records will provide a systematic but selective, holistic and context-sensitive description of individuals within the study. Presented thematically according to the domains of the interview guide, they were used to describe the patients experience as reported in the baseline and follow-up interviews. In this sense they represented a storytelling approach to reporting data as development over time was also described.¹⁸⁵ On a simpler descriptive level, illustrative examples of baseline and follow-up comparison of SEIQoL cues (with corresponding levels and weights) have been presented.²⁰¹

Credibility of the qualitative analysis: The use of memos, coding strategies and case study analysis contribute to sustaining the overall credibility of the qualitative analysis. More formally, the issue of credibility was addressed in a number of ways. Firstly, a small number of interview transcripts were reviewed initially in the process of category development. Emergent categories were tested against new data from remaining transcripts. This method of testing out provisional hypotheses on further cases is termed *constant comparison*.²⁰⁹ A second, but related process, was employed - that

of *deviant-case analysis* whereby data elements discrepant with the developing classification schema are identified. The schema was modified to incorporate such cases.^{185 206 209}

Although a variety of methods were used to generate data within the interview (e.g. standardised health status, individualised quality of life, semi-structured qualitative questioning and probing) they may all be susceptible to self-report bias. Therefore, *triangulation* by use of different data collection approaches helped account for the limitations of any particular method.²⁰⁶ Drawing together these sources serves to enhance the credibility of general conclusions.²¹⁰

5.8 Summary of analytic plan

The overall analytic plan for the study interviews is shown in table 5.8.1. It summarises the aims and objectives represented by each analytic component; whether the approach is broadly qualitative or quantitative; and indicates within which chapter results are presented and discussed.

Table 5.8.1 Response shift study: aims, objectives and output map

Analysis	Research aim / objectives	Analytic approach		Chapter
		Qualitative	Quantitative	
a	Describe and classify the nature of individual quality of life domains nominated by patients with ongoing knee problems	✓		7
b	Determine whether patients with knee problems experience a response shift in their self-reported quality of life over a period of six months	✓		8
c	Explore evidence of re-conceptualisation response shift	✓		8
d	Explore evidence of re-prioritisation response shift	✓		8
e	Explore mechanisms that may affect response shift (behavioural, cognitive and affective processes)	✓		8
f	Evaluate the impact of response shift upon the responsiveness of two methods of evaluating status (EQ-5D and SEIQoL)		✓	9
g	Explore evidence of, and associations with re-calibration response shift		✓	9
h	Explore the relationship between quality of life domain level and its perceived importance and its impact upon reported quality of life	✓	✓	10
i	Evaluate the use of SEIQoL-DW within a quantitative analysis of response shift		✓	11
j	Assess the impact of delay in diagnosis and treatment upon quality of life	✓		12

5.9 Summary

This chapter has presented the rationale and methods for the studies that form the basis of the response shift theme of the thesis. It has described how the thesis studies have been partly set within the clinical research framework implemented for The DAMASK study. The DAMASK framework was designed in part with the thesis studies in mind, but has been extended where necessary to accommodate the requirements of the thesis studies. The chapter has described the design of the thesis studies including the development of the interview schedules, sample specification, choice of outcome measures and broad qualitative analysis plan (including theoretical justification). The component studies have been delineated and clear objectives for each described. The summary analysis plan describes the quantitative or qualitative emphasis of each component study and shows within which chapter results are presented. The choice of method (principally qualitative assessment of individualised quality of life data) has been selected due to the relative paucity of such approaches in assessing response shift identified in the literature review presented in chapter 4. The method addresses both re-prioritisation and re-conceptualisation response shift. Nevertheless, the interview framework also incorporates a more conventional retrospective pretest-posttest study to provide complementary response shift assessment (specifically re-calibration). The latter is presented in chapter 9.

Preface to chapter 6

The current chapter sets the framework for all the remaining empirical work presented in the thesis. The following chapter presents a largely descriptive overview of the study sample obtained and a description of the process of conducting the baseline and follow-up interviews. It is intended as a brief summary and starting point for the results of the component response shift studies.

Chapter 6: Response shift interview study: overview and descriptive results

***Abstract:** This chapter summarises the recruitment and the characteristics of the interview study sample. Patients with suspected or confirmed internal derangement of the knee were recruited from the departments of radiology and orthopaedics of two local NHS hospitals. 39 patients were successfully interviewed at baseline and all but three of these people completed a follow-up interview approximately seven months later. The interviewed sample reflected the purposive sampling strategy which maximised variation according to gender, age and stage of clinical management. Summary statistics for key quantitative outcome measures are described. This includes baseline, follow-up and 'thentest' (retrospective rating of baseline state) scores for the health utility measure EQ-5D. Mean index scores for the individualised quality of life measure, SEIQoL-DW are also reported. No statistically significant differences were found on either EQ-5D or SEIQoL-DW between patients in the arthroscopy group and waiting list group when using prospectively derived scores. The chapter as a whole provides an introduction to the component studies presented in chapters seven to twelve.*

6.1 Summary of patients approached, consented and interviewed

Eligible patients for the waiting list group were identified through general practitioner referral letters to the Department of Orthopaedics (Llandough Hospital) and general practitioner requests for MRI investigation to the Department of Radiology (University Hospital of Wales). For patients referred to orthopaedics, details were collected from 22nd November 2002 until 15th January 2003. For patients referred to radiology, details were collected from 6th December 2002 until 24th February 2003.

Eligible patients for the arthroscopy group were identified through patient details recorded in orthopaedic consultant operating diaries (Llandough

Hospital). For such patients, details were collected from 11th November 2002 until 22nd January 2003.

Between 27th November 2002 and 6th March 2003, 106 patients were approached for consent by letter (47 arthroscopy patients; and 59 waiting list patients recently referred by their GP to orthopaedics, or for investigation by MRI). In the arthroscopy group, 37 patients had 'arthroscopy' or 'endoscopy' recorded in their pre-operative notes, the remainder had anterior cruciate ligament work identified (see table 6.1.1). In the waiting list group, 25 patients referred to the Department of Radiology had the meniscus or ligament explicitly mentioned in the referral letter. All remaining radiology patients had symptoms or signs suggestive of internal derangement of the knee mentioned in the letter (pain; locking; giving way; or swelling), apart from one patient who simply had an injury reported. Seventeen waiting list patients referred to the Department of Orthopaedics had the meniscus or ligament explicitly mentioned in the referral letter. In all but one of the remaining eight, signs or symptoms consistent with IDK were mentioned in the referral letter.

Table 6.1.1 Overview of patients approached for study

		Clinical details: from pre-operative notes		Clinical details: from referral letter		
		<i>Arthroscopy mentioned</i>	<i>ACL mentioned¹</i>	<i>Meniscus / ligament mentioned</i>	<i>Signs or symptoms consistent with IDK²</i>	<i>Other</i>
Orthopaedics						
<i>Arthroscopy group</i>	(n=47)	37	10	-	-	-
<i>Waiting list group: referred for consultant assessment</i>	(n=25)	-	-	17	7	1
Radiology						
<i>Waiting list group: referred for MRI</i>	(n=34)	-	-	25	8	1

1 *If arthroscopy not otherwise explicitly mentioned*

2 *If meniscus or ligament not otherwise explicitly mentioned*

Consent forms were signed and returned to the researcher (MR) by 45 patients (42%). Baseline interviews were arranged and completed with 39 patients. The remaining six patients who provided consent were not interviewed for a variety of reasons including subsequent inability to be contacted during the study period, despite repeated attempts. The numbers of patients approached; providing consent; and completing a baseline interview within each of the study groups is shown in table 6.1.2. The total number of patients completing the first interview is also shown by study sub-group in table 6.1.3. The proportions of approached patients completing and not completing interviews by each stratifying factor (age, gender and stage of management) were compared using chi-square tests. There were no significant differences found between men and women; between older or younger patients; and between listed and referred patients in those completing or not completing interviews.

Table 6.1.2 Number of patients approached, providing consent and completing baseline interview

		Stage of management		
		<i>Arthroscopy</i>	<i>Waiting list</i>	
Gender	<i>Male</i>	15 / 5 / 4*	15 / 5 / 5	<40 Age
		11 / 5 / 5	15 / 6 / 6	40+
	<i>Female</i>	9 / 5 / 4	13 / 5 / 3	<40
		12 / 7 / 7	16 / 7 / 5	40+

*Counts are: Approached / Consent provided / Baseline interview completed

Table 6.1.3 Number of patients interviewed at baseline by sample sub-group

Sample sub-group	Group	Total (n)
Stage of clinical management	<i>Arthroscopy</i>	20
	<i>Waiting list</i>	19
Gender	<i>Male</i>	20
	<i>Female</i>	19
Age	<40	16
	40+	23

6.2 Completion of baseline and follow-up interviews

The baseline interviews lasted approximately 45 minutes, whilst the follow-up interviews were slightly longer - with a few lasting just over an hour. Three patients interviewed at baseline were not followed-up by a second interview. One was not at home on any of three planned appointments, one was not contactable despite repeated telephone calls and it proved impossible to arrange an appointment with the third during the study period. For the latter, a postal version of the standardised interview schedule was despatched but was not returned. Most interviews were conducted as planned at the interviewees' home, although a few were conducted either at the interviewees place of employment or in offices of the academic Department of General Practice. Most interviews were conducted on a one-to-one basis, although for a small number of interviews partners or young children were also present.

Although most interviewees appeared to have few problems completing the interview, including the SEIQoL-DW and the EQ-5D components, a few patients did appear to have some difficulties. These included the assignment of weightings to their nominated cues and confusing rating of cue level with weighting. Not all interviewees were able to nominate five cues, and some were very focused upon their knee problem when completing the SEIQoL-DW. The order of the assessments was changed on a few occasions to allow the interviewee to express their feelings about the impact of their knee problem before addressing broader quality of life issues. A number of interviewees also had problems responding to the EQ-5D items – especially choosing between the broad response categories provided by the measure. Such respondents were generally encouraged to choose the closest appropriate option, but this was still problematic for those, for example, who had been experiencing intermittent difficulties. At baseline, one interviewee, for whom English was not his first language, appeared suspicious of the true aim of the study. After much discussion to reassure the interviewee, the interview proceeded, but the formal SEIQoL-DW assessment was omitted (as was also the case in the follow-up interview).

The mean follow-up interval for the 36 patients completing a second interview was 217 days (approximately seven months). The minimum interval was 185 days and the longest was 268 days. The distribution of follow-up intervals is shown in figure 6.2.1. Most of the baseline interviews were conducted during the winter-spring months and consequently most of the follow-up interviews were conducted during the summer-autumn months (figure 6.2.2).

Figure 6.2.1 Distribution of follow-up intervals

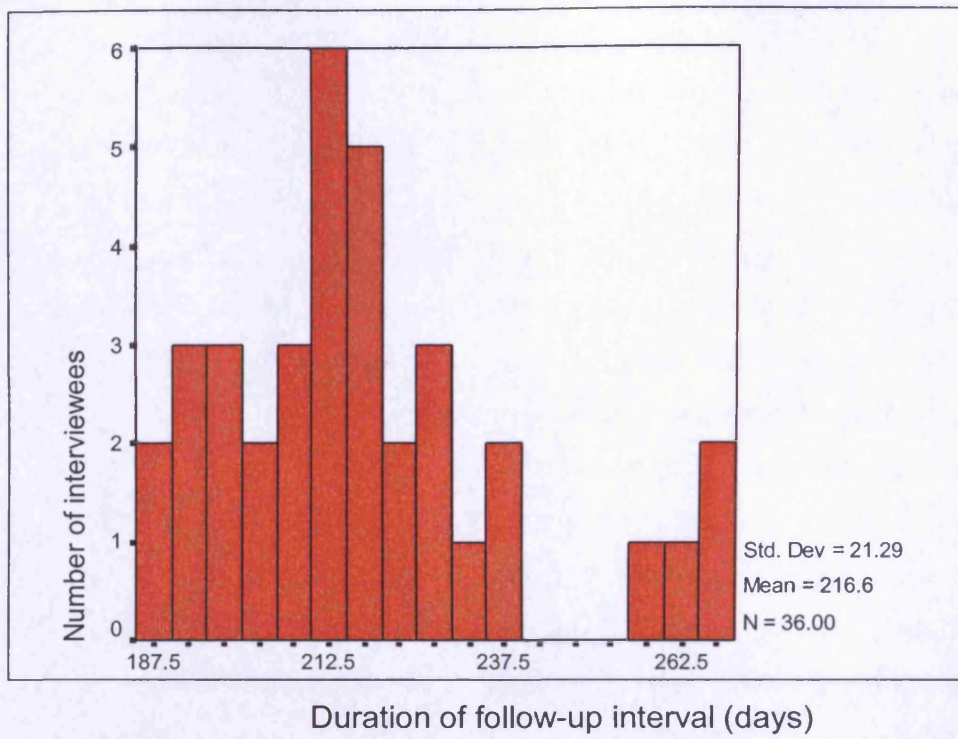
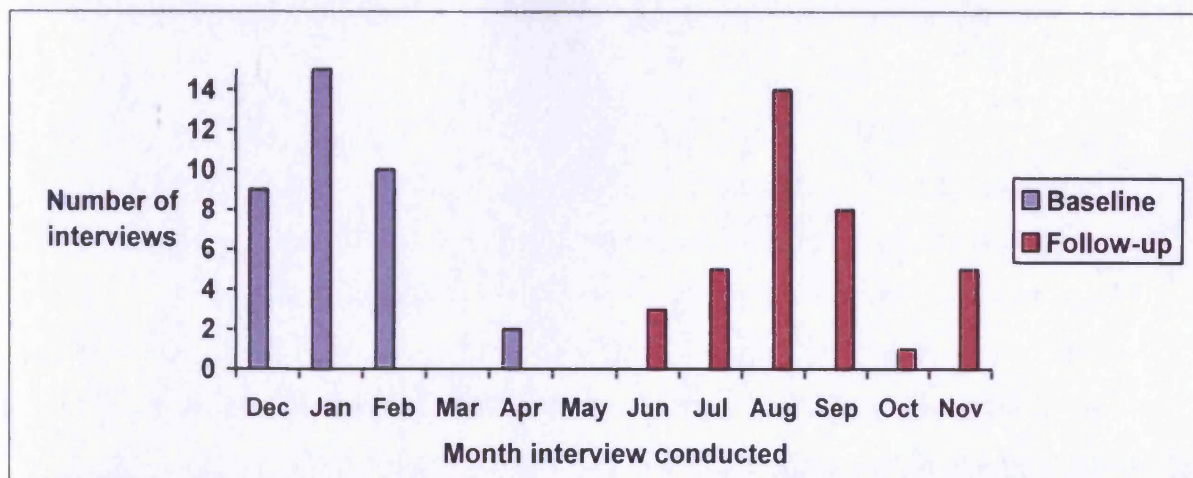


Figure 6.2.2 Timing of baseline and follow-up interviews (for the 36 patients completing paired interviews)



6.3 Description of sample

The mean age of the 16 patients in the lower age group was 30.4 years (table 6.3.1), whilst that of the 23 patients in the upper age group was 50.4 years. The youngest patient was 19 and the oldest 64 years of age. There were 30 patients identified from the Department of Orthopaedics (across the three clinical collaborators: CW, RMJ and RW) and nine from the Department of Radiology.

Table 6.3.1 Summary descriptive data for study sample (n=39)

Variable	Level	Mean (sd)	
Age	<40	30.4 (6.6)	
	40+	50.4 (6.9)	
LOT-R (dispositional optimism) ¹		14.1 (4.27)	
Collaborating consultant (identified by initials)	Orthopaedics	CW	13 (33.3%)
		RMJ	7 (17.9%)
		RW	10 (25.6%)
	Radiology	KL	9 (23.1%)

¹ For those completing follow-up interview

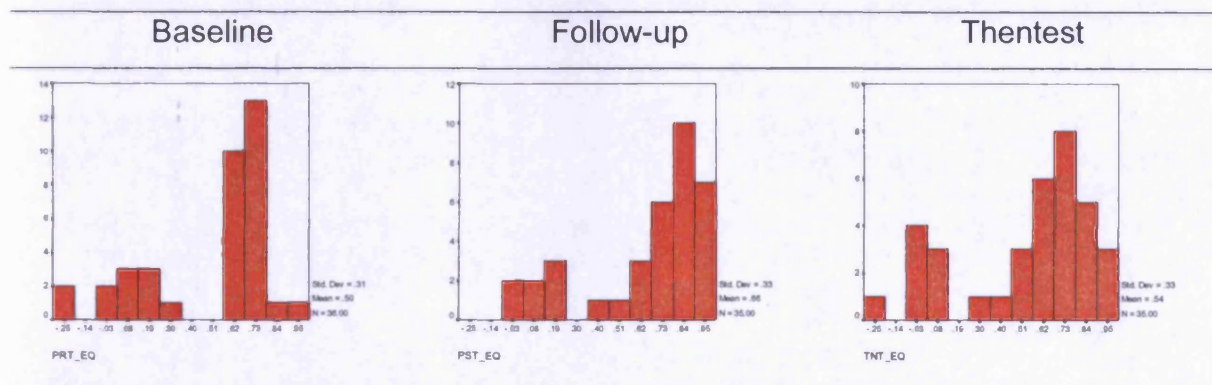
The mean dispositional optimism score on the LOT-R was 14.1 (sd: 4.27). Higher scores represent greater optimism. This compares with norms for US college students of 14.33 (sd: 4.28) and coronary artery bypass patients of 15.16 (sd: 4.05).²⁰⁵

Seventeen of the 20 arthroscopy patients reported at follow-up that they had undergone surgery. For one patient (#14), the consultant had decided to monitor and review his case. A second patient (#30) had initially opted to defer her operation, but reported at follow-up that she had requested to be put back on the list for surgery. A third patient (#28) reported at follow-up that she was now awaiting a privately funded total knee replacement. Of the 16 waiting list patients completing a follow-up interview, none reported receipt of surgical intervention in the interim study period.

Outcome measures: summary of sample scores and comparison of arthroscopy and waiting list patients

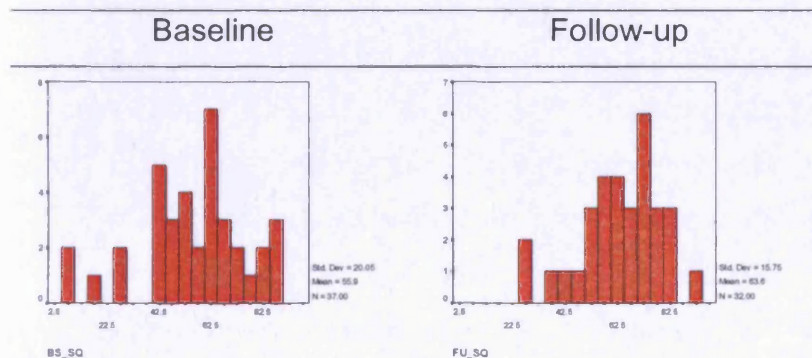
The distributions of EQ-5D scores from baseline, follow-up and using the retrospective rating of baseline status ('thentest') were all negatively skewed (figure 6.3.1). In contrast, the distributions of SEIQoL-DW scores at baseline and follow-up were much less skewed, especially the former (figure 6.3.2). Summary statistics for the whole sample are shown in table 6.3.2. The median baseline score for EQ-5D was 0.62, with a median score at follow-up of 0.73. The median 'thentest' EQ-5D score was 0.62. The mean baseline SEIQoL-DW score for the study sample was 55.9 (sd: 20.05) and at follow-up was 63.6 (sd: 15.75). Summary EQ-5D and SEIQoL-DW statistics for arthroscopy and waiting list patients are shown separately in table 6.3.3. Differences between the listed and referred groups on each of the quality of life measures at both baseline and follow-up assessments were tested using Mann-Whitney U-test. No statistically significant differences were found.

Figure 6.3.1 Distribution of baseline, follow-up and thentest EQ-5D scores



Higher EQ-5D scores represent better health

Figure 6.3.2 Distribution of baseline and follow-up SEIQoL-DW scores



Higher SEIQoL-DW scores represent better quality of life

Table 6.3.2 Outcome measure scores for all patients at baseline and follow-up

Quality of life score	Median (iqr)	Mean (sd)
EQ-5D		
Baseline	0.62 (0.5)	0.50 (0.31)
Follow-up		
Posttest	0.73 (0.36)	0.66 (0.33)
Thentest	0.62 (0.47)	0.54 (0.33)
SEIQoL-DW		
Baseline	59.92 (24.8)	55.9 (20.05)
Follow-up	64.63 (20.09)	63.6 (15.75)

Table 6.3.3 Quality of life scores for listed and referred patients at baseline and follow-up

Outcome measure score	Study group			
	<i>Athroscopy</i>		<i>Waiting list</i>	
	<i>Median (iqr)</i>	<i>Mean (sd)</i>	<i>Median (iqr)</i>	<i>Mean (sd)</i>
EQ-5D				
<i>Baseline</i>	0.62 (0.6)	0.41 (0.36)	0.62 (0.1)	0.60 (0.22)
<i>Follow-up</i>				
Posttest	0.73 (0.48)	0.66 (0.35)	0.77 (0.46)	0.66 (0.32)
Thentest	0.62 (0.44)	0.54 (0.36)	0.62 (0.59)	0.54 (0.31)
SEIQoL-DW				
<i>Baseline</i>	60.08 (23.46)	55.25 (24.07)	53.20 (27.87)	56.63 (15.40)
<i>Follow-up</i>	63.67 (18.98)	62.80 (15.86)	69.06 (22.55)	64.48 (16.12)

6.4 Discussion

Although discussion about data from the response shift interview studies is mainly presented where appropriate in subsequent chapters, two issues are worth addressing here. Firstly, the qualitative sampling strategy had sought to maximise variation by ensuring a balance of participants within each main stratifying variable (age, gender and stage of clinical management). Although the baseline sample had one less participant than planned, the requirement for adequate balance within and across strata was satisfied. Some failure to follow-up all initially interviewed patients had been expected, and this was reflected in the original sample size chosen. The final total of 75 completed interviews (including 36 paired assessments) was considered sufficient for providing reasonable coverage of the phenomenon being explored.¹⁸⁵

Secondly, although the planned interval between assessments was six months, it was actually a month longer on average. Reasons for follow-up being later than expected included logistical and practical difficulties in contacting patients and arranging a second interview. However, the planned interval was chosen to allow those patients listed for arthroscopy time to experience a change in their physical functioning (i.e. they were expected to improve). In contrast those patients who had been recently referred were not expected to improve during the follow-up period without surgical intervention. Therefore, the delay in follow-up was unlikely affect assessment of either

group. For the former, it simply allowed them longer to recover. For the latter group, the waiting times for initial orthopaedic consultation (over two years) were far in excess of the seven month interval from baseline to follow-up interval.

6.5 Summary

A sample of patients reflecting the diversity indicated by the purposive sampling strategy was successfully recruited and interviewed on two occasions. The interviews and their constituent assessments appeared acceptable and were largely completed successfully. Descriptive statistics for the main quantitative outcome measures (SEIQoL index score and EQ-5D) are presented. Although no apparent difference between waiting list and arthroscopy patients at either baseline or follow-up was found, further analysis is presented in chapter 9.

Preface to chapter 7

The following empirical chapters each address separate analytic aims within the context of the overall response shift themed studies. The first of these addresses the nature of quality of life as described by interviewees in response to the SEIQoL-DW assessment. It thus seeks to provide an individualised perspective to quality of life experience for patients with knee problems. The analysis provides a baseline from which evidence of re-prioritisation and re-conceptualisation response shift is explored in chapter 8.

Chapter 7 A thematic analysis of quality of life content

Abstract: *The increasing profile of quality of life assessment within health research, and developments in its conceptualisation and assessment are described in the introduction to this chapter. The introduction also includes a focus upon the use of individualised approaches to assessment. Inductive content analysis used data from both baseline and follow-up interviews. Although it mainly drew upon data generated by use of SEIQoL-DW, it also included data from other sections of the interview. Analysis was facilitated by the use of memos, categorising strategies and contextualising strategies. Analysis resulted in the construction of a thematic framework of quality of life areas. Twelve main areas are described and include commonly assessed domains such as health; work; family; and finance; as well as less frequently considered areas (for example, spiritual and religious content). Each of the main themes was further sub-divided into categories representing different aspects of interest. For example, the health domain was represented by sub-themes of enabling health; physical, mental and functional health; and positive and negative health. The nature of each theme and how they relate to one another are illustrated with extracts from the interviews.*

The derived framework contributes to the response shift theme of this thesis in two ways. Firstly, it provides an insight into the complex and idiosyncratic way in which quality of life was experienced by this sample of patients, and moves beyond simple physical and functional based assessment. Secondly, it provides a baseline for the detailed assessment of cue changes across interviews presented in chapter eight. Specifically, it facilitates response shift analysis which does not end with comparison of cue label, but progresses towards a greater consideration of change that may occur within, as well as between major content areas.

7.1 Introduction

The increasing profile of quality of life assessment

Quality of life has become a commonly used term, both in professional and lay circles, but what is understood by the term, and of what importance is it in healthcare research and practice? In 1996, Testa and Simonson charted the increasing importance of quality of life issues in health care research and practice since the World Health Organisation's original expanded definition of health as a state of complete physical, mental and social well-being.^{14 211} Their Medline review revealed a six-fold increase in the use of the key word 'quality of life' between 1973 and 1996. Similarly, in 2002, Garratt and colleagues described in their bibliographic review an increase in reports of the development and evaluation of patient reported quality of life measures from 144 in 1990 to 650 in 1999.²¹² Three medical specialties accounted for nearly a third of all 3921 records found (rheumatology and musculoskeletal medicine, cancer and older people), although only 65 were found for orthopaedics.

The increasing profile of quality of life within health research partly reflects a continuing imperative to quantify healthcare outcomes to justify resource allocation and restriction, and to satisfy the rigour of evidence-based medicine.^{198 213} Other catalysts include a clinical shift towards viewing patients more holistically and as means for sharing treatment decisions with patients to enhance their autonomy.^{197 198 214} Perhaps most fundamentally, the main objective of a healthcare system may be viewed as increasing the quality of life of its served population.²¹⁵

Conceptualisations of quality of life

Day and Sankey noted an historical development in approaches to quality of life, with an initial focus upon economic approaches.²¹⁶ This shifted to an interest in societal indicators and (whilst these are still utilised) subsequently

towards a focus upon psychological and subjective indicators.¹ More recently, 'gap' theories have been proposed which seek to determine the difference between current experience of life and individual expectations or standards.²¹⁹ Farquhar reviewed definitions of quality of life and subsequently proposed a taxonomy in which she distinguished between those that were professional and lay.²²⁰ Amongst the former are all-encompassing *global* definitions which are too general to be operationalised and describe little of quality of life content. A second set of definitions distinguish quality of life *components*, or specify characteristics considered essential for evaluation. Conducive to operationalisation, component definitions were also sorted according to whether they were specific or not to a research topic. Thirdly, focused definitions make reference to either one or a small number of components (e.g. health and health-related quality of life) either explicitly or implicitly. A fourth group of combined definitions seek to integrate global definitions with specific components. In contrast to expert or professional definitions, lay definitions are an attempt to determine subjective meanings drawn from specific populations.

There remains however, a lack of conceptual clarity and thus, consensus as to what quality of life should refer.^{213 221-224} Indeed, Aristotle argued that most people agreed on the primary importance to humans of happiness (eudaimonia²) but that what this meant for any one individual varies.²²⁶ Two separate concept analyses of definitions of the quality of life construct were conducted by Meeberg in 1993 and Haas in 1999.^{227 228} The former identified four repeatedly appearing attributes: (i) a feeling of satisfaction with one's life; (ii) the mental capacity to evaluate one's life; (iii) an acceptable self-assessed

¹ Campbell cited an unattributed quote referring to President Lyndon Johnson's Great society using the phrase 'quality of their lives'.²¹⁷ Campbell A. *The sense of well-being in America: Recent patterns and trends*. New York: McGraw-Hill, 1981. However, in his Great Society speech at the University of Michigan, 1964 Johnson specifically describes the Great Society as 'a place where men are more concerned with the quality of their goals than the quantity of their goods'.²¹⁸ Johnson LB. Great Society Speech, University of Michigan, 1964.

² Aristotle used the noun eudaimonia which is generally translated as happiness but has alternatively been considered to represent a fulfilled life. For example, Carol Ryff has argued for the concept of eudaimonic well-being as the realisation of one's true potential (as distinct from hedonistic well-being).²²⁵ Ryff C, Singer BH, Love GD. Positive health: connecting well-being with biology. *Philosophical transactions of The Royal Society of London. Series B: Biological Sciences* 2004;359:1383-1394.

state of physical, mental, social and emotional health; and (iv) an objective assessment by another that one's living conditions are adequate and not life-threatening.²²⁷ Haas also described four features, the first being a multidimensional and dynamic nature which thus varies with context.²²⁸ Quality of life is also an appraisal or evaluation, and is subjective in nature. Finally, Haas delineated an objective component to quality of life, with behaviour, functioning and environment being used as indicators of quality of life. Although both reviews describe self-evaluation, this seems at odds with the inclusion of an objective perspective. Furthermore, concept definitions resulting from each review are not theoretically derived, rather the result of a synthesis of existing definitions.

Assessing quality of life

Hunt described three basic approaches to quality of life assessment: health-related quality of life (HRQoL); quality of life years (QALY); and conceptual models.²¹³ HRQoL approaches developed originally to assess health status assume that health status impacts upon quality of life. Hunt noted that such measures don't account for individual weighting of domains and ignore variation in coping. The latter criticisms were also applied to QALYs by Hunt. Included amongst the conceptual models are gap theories (e.g. described by Calman) and individual cognitive approaches (e.g. SEIQoL).^{188 219} In general, Hunt criticised the lack of empirical data used to inform conceptual models which therefore, remain inadequately refined. Specifically, SEIQoL was criticised for assuming that aspects of the respondent's 'condition' will be related to quality of life. Whilst SEIQoL doesn't ask respondents to reference their medical condition (unlike, for example, the Patient Generated Index), it probably does address determinants of quality of life, rather than assess it directly.^{59 186 192}

Whilst there is some agreement that quality of life is multi-dimensional and subjective there are conceptually divergent approaches to its assessment.²¹⁵ A common approach to assessment is the use of standardised multi-dimensional measures reflecting standard needs.²²⁹ Critical within this approach is the selection of domain content which may include physical and

health status; functional ability; psychological status and well-being; social interactions and economic status. However, what is important at one point in time may vary between individuals, and relevant content may vary over time within individuals.²³⁰ Identifying what should be included within a measure is a critical psychometric concern.^{213 231} After more basic human needs (e.g. food) have been satisfied, agreement about which 'wants' contribute to quality of life is harder to achieve. Even for content for which there is a shared interest the relative importance to overall quality of life may vary between individuals and therefore misrepresent the hierarchy that exists in real life.²¹⁵

Individualised approaches to assessment

Perceived weaknesses of the standardised nomothetic approach have led some to argue for an individual ideographic approach to quality of life assessment.^{215 231 232} Such criticisms have included the presentation of predetermined content domains which may not reflect the interests of any one individual.²³³ Secondly, it has been argued that quality of life must be rooted in existentialism and reflect individual values and the dynamics of human experience.²¹³ The inability of normative approaches to reflect the relative importance attached to different aspects of life has been a common criticism.^{198 221 227 232} The imposition of an predetermined frame of reference which is universally weighted challenges the relevance of outcome score for any one individual.²⁰¹

A related concern is the implied normality or desirability of certain states and the narrow juxtaposition of good quality of life with good health status.²¹³ Furthermore, normative quality of life scores may have no intrinsic meaning, making interpretation more difficult and approaches (e.g. qualitative) which explore decision-making or appraisal may be more instructive.^{95 234} To address the issues of individual relevance measures such as SEIQoL utilise a working definition of quality of life as something that an individual determines it to be.²³³ The approach recognises the dynamic nature of quality of life and that the relative importance of component domains may vary. A further assumption underlying the SEIQoL is that quality of life is equivalent to the

aggregated satisfaction with a small number of critical life domains, which are themselves individually weighted.²²⁹

Clinical applications of quality of life

Despite scientific doubts about the current stage of theoretical development, and moral concerns about its premature application, quality of life assessment is increasingly finding a role in clinical practice and research.²¹³ Perhaps not least this is because 'it places the patient at the heart of the therapeutic process'.²³⁵ Quality of life may be the primary outcome in a clinical study and its place amongst other clinical variables in clinical trials and other evaluative studies is well-established.^{136 221 236 237} Quality of life may not only be a measure of treatment outcome but also a means of sharing clinical decision making with patients; auditing care; monitoring progress; planning treatment; prioritising patients for treatment; and individualising care.^{197 237-241} It has also been recognised that with improving technology able to extend life, quantity should be balanced by life quality.²²⁸ Eliciting individual quality of life concerns may facilitate information exchange within consultations, functioning as a clinical intervention.^{237 242} SEIQoL-DW has been promoted as such in palliative care by facilitating doctor-patient communication; guiding and focusing interventions; and helping to develop coping strategies.²⁴³ However, evidence for the effectiveness of interventions based upon the use of outcome measures is currently equivocal.²⁴⁴

Considering quality of life may take on particular importance within certain patient groups, such as those with cancer, where the primary aim of treatment is improving life quality by curing disease or alleviating symptoms.²³⁸

Balancing treatment benefits with treatment toxicity and other potentially important side-effects is particularly relevant in oncology and palliative care.²³⁵

²³⁸ Indeed, quality of life is integral to the definition of palliative care adopted by the World Health Organization, and at the inauguration of the medical speciality of palliative medicine in the UK.^{245 246} Thus, Schwartz and Sprangers delineated four achievements of quality of life research in oncology: the assessment of treatment outcome and the qualification of survival; the assessment of late problems in long-term survivors; the

prediction of mortality; and as a facilitator of communication between patients and clinician.^{160 237}

Assessing quality of life is of increasing relevance for patients living with chronic conditions such as rheumatoid arthritis and diabetes, where major differences in evaluative studies (for example, in survival rates) are unlikely.¹⁵⁹²⁴⁷ Objective indicators alone are increasingly viewed as insufficient in evaluating patient experience.²³⁵ There may not be a simple relationship between traditionally assessed clinical signs and patients' evaluation of experience. A more holistic assessment of individual functioning may be more relevant for the patient.¹⁵⁹ However, the clinical use of quality of life measures may present challenges, not least the time required for their application.²⁴⁸

Study aims

The aim of this part of the study was to firstly describe the nature of quality of life content expressed by patients within the study sample. This was conducted using an individualised approach (modified SEIQoL-DW) which enabled respondents to discuss what was of particular importance to them. This allowed an exploration of how quality of life was understood and how it was constructed by individuals (for example, the interaction and inter-relationship between different domain content). The second aim of the analysis was to provide a base against which changes in quality of life prioritisation and conceptualisation could be assessed.

7.2 Methods

The process of sample identification, approach and interview was described in chapter five. Similarly the broad qualitative analytic approach of conceptual ordering was described in the same chapter. Data analysis employed the creation and use of memos; categorising strategies; and contextualising strategies.²⁰⁶

The purpose of the analysis was to produce a descriptive thematic framework. The framework themes were not intended to be wholly mutually exclusive, and there was no intention for data to be precisely coded under one or other theme (a fuller consideration of this issue is presented in chapter eleven). Whilst sampling for the study was not intended to achieve category saturation (i.e. by conducting further interviews until no new categories arose), it was nevertheless predicted that most major themes would arise within the chosen sampling frame.²⁴⁹

Data from both baseline and follow-up interviews were used. Whilst descriptions of quality of life cue content were most likely to be found in the application of SEIQoL-DW, data from other parts of the interview were considered if relevant, (for example, if the respondent mentioned quality of life when subsequently discussing their knee problem). The process of identifying themes was inductive, although the analyst (MR) was aware of other categorisations of similar data generated from SEIQoL. Content analysis represented the first of the grouping areas identified a priori (described in chapter five).²⁰⁶

Explicit in the analytic process was that the resultant framework would represent the analyst's perspective of the internal homogeneity of themes and of how they related together (external homogeneity).²⁵⁰ Analysis therefore, worked from the respondents' own extended description of quality of life content but was not restricted, for example, by the respondents labelling of cue content.

In the interview, and when analysing the data to construct the framework, two questions were considered. Firstly, what function was served by the attribute described by an individual? For example, if 'work' was provided as a cue label, what aspects of the global phenomenon of work was being described? Secondly, how did the described attribute contribute to quality of life? For many domains, it was clear to the interviewee how it affected quality of life. However, in some instances a clear description was not forthcoming from the interviewee and was more difficult to infer from the associated conversation.

7.3 Results

Twelve main themes were identified, and are shown in table 7.3.1 below. Also described are the key sub-themes identified for each main domain. Whilst themes are numbered, no hierarchy or weighting is implied. In the results section, each theme and sub-theme are described along with a description of the quality of life content included within it. To this end, typical quotes are provided to illustrate themes and sub-themes. For each interview extract provided a study record number is used to identify the interviewee and from which interview the extract was taken (i.e. I=baseline; II=follow-up). Respondent details are provided below each extract to identify their gender; age; stage of clinical management (waiting list or listed for arthroscopy); and occupation recorded at baseline. For the sake of brevity, respondents are occasionally referenced without an accompanying quote.

Table 7.3.1 Categorisation of Quality of Life Domains: Major headings

Quality of life domain		Key sub-themes within each domain	
I	Work	i)	Ability to work
		ii)	Security at work
		iii)	Professional interest
		iv)	Stress
		v)	Income generation
		vi)	Social component
II	Family	i)	Stability
		ii)	Immediate and extended family
		iii)	Family role
		iv)	Family activities
		v)	Pets
III	Money	i)	Basic provision and debt
		ii)	Enabling finance
		iii)	Long-term security
IV	Social life and friends	i)	Social activities
		ii)	Level of organisation
		iii)	Following sport
		iv)	Friends
V	Health	i)	Enabling health
		ii)	Physical, mental and functional health
		iii)	Negative and positive health
VI	Physical activities	i)	Sporting activities
		ii)	Fitness activities
		iii)	General physical activities
VII	Leisure & pastimes	i)	Gardening
		ii)	Hobbies
		iii)	General leisure
VIII	Personal environment	i)	Home ownership
		ii)	Emotional appraisal of home
		iii)	Home maintenance
		iv)	Neighbourhood
IX	Psychological constructs	i)	Independence
		ii)	Achievement
		iii)	Adventure
		iv)	Relaxation
		v)	Stability
		vi)	Social harmony and peace
		vii)	Getting on with life
X	Education	i)	Establishing a career
		ii)	Education for life
XI	Spiritual and religious	i)	Personal faith
		ii)	Religious commitment
		iii)	The Church community
XII	Holidays	i)	Family and social functions
		ii)	Escape and distraction
		iii)	Mental and physical health benefits
		iv)	Tourism

I Work

A variety of attributes related to work was included under this general header. This included aspects inherent to the content of the job itself as well as other attributes of work and the work setting. Sub-headings included:

i) Ability to work: This related to whether or not the interviewee was able to work, usually in reference to their knee complaint, but also other health problems. Some patients were either currently or recently unable to work. The consequences for their income and broader well-being were expressed in this context (the interdependence between separately nominated quality of life cues is described later). At least one interviewee was self-employed and had mentioned this concern in particular. The ability to conduct a full range of activities was also included here, although as an interview topic this was commonly discussed in subsequent parts of the interview rather than being expressed as a quality of life area.

ii) Security at work: This included 'security' related to the knee problem and its impact upon the patient's work as well as other issues. An example of the latter is a patient who was temporarily upon secondment from their normal job. Although he had a secure job to return to, there was uncertainty involved in returning from his secondment with what changes he would be confronted with (II#11). Security in this sense entailed a sense of familiarity and continuity. A second patient referred to the security of knowing what their job was and how to perform within it (II#7). This concept of 'security' addressed a sense of professional role considerations.

Work security reflected for some (or enabled) general stability in their life (e.g. I#3) and being able to make firm plans for the future. Work security enabled control. Whilst work security and financial security are obviously closely linked, some respondents separated them out as distinct but inter-related considerations, (for example, I#32).

iii) Professional interest: A number of related concepts were included under this sub-heading, including work providing interest and stimulation, enhancing mood, satisfaction, enjoyment, opportunities for self-actualising, personal and professional development and maintaining a life-role. Many of the interviewees expressing sentiments within this category were from recognised 'professions' (e.g. teacher, engineer). One respondent, a secondary school teacher, referred to the professional interest of his current work being an indicator of his personal happiness:

I#11: Yes, of this nature, so that would be another indicator for my personal happiness to me.

(Male / 44 years old / referred / teacher)

Furthermore, work provided the individual with a focus or purpose within their life (e.g. I#14). In one sense this may be regarded as means of avoiding a negative (i.e. a state of aimlessness), rather than a fundamentally positive state. A more positively framed expression of this aspect was the opportunity for work to enable self-actualising behaviour (e.g. to succeed, as described by patient I#13) or self-esteem:

I#2: No, it gives you self esteem as well because of challenges ...

(Male / 40 years old / listed / civil engineer)

Work was a source of stimulation with the stress and pressure of work regarded overall as a desirable and positive experience. This may be recognised in facets such as working to deadlines, variety within the job, and achieving a satisfactory balance between excess and insufficient stress. The following comment illustrates this:

I#32: Well, the challenge. You know it's um, there's a certain amount of stress which is absolutely necessary but it's positive stress. You wouldn't want negative stress in your job 'cos that's just counter-productive. If it's positive stress, it's the challenge of things changing, moving, continually being pushed. I couldn't think of anything worse than actually going to work and it's the same thing every day.

(Male / 46 years old / referred / IT project manager)

Work, therefore, provided intellectual stimulation and with it the prospect of achievement within the work role. Work provided a social role for individuals which, as the following quote shows may be in addition to other roles in life or even a replacement for 'lost' roles:

I#39: If you're a certain person, your role is to bring your children up and bring them up as best you can and then when they grow up and leave home you've lost that kind of role. So you replace it with another role and for me it was to going to work full-time. You know, a tidy job as opposed to part-time jobs, so you give it 120% or 130%.
(Female / 50 years old / referred / administrator)

Two further examples:

I#2: You get a genuine kick out of work sometimes don't you?
(Male / 40 years old / listed / civil engineer)

I#38: I could never stay at home and just potter around in the house. If I did that I'd want to go out all the time and I'd want to spend money and you know. I did go to friend's houses and things but that was never enough. I felt I needed something to get on with, get the brain working .
(Female / 47 years old / referred / teacher)

Most comments related to professional interest were positively framed but occasionally stated due to a current absence of stimulation or interest. That the stimulation and pressure within work was a double-edged sword (providing pleasure and stress) was also referred to (e.g. I#5).

Work was not always mentioned as being important to an individual's overall quality of life, and for those working full-time the mention of work was perhaps noticeable by its absence. Indeed, some explicitly stated that it was not important, and favoured mentioning other concerns such as 'family'. Nevertheless, the necessity to be happy in one's work was discussed - the very amount of time spent working was reason enough to see this as a necessary source of satisfaction and stimulation:

I#32: I mean, part of the reason is that um, in order for you to be content and be happy, it's all very well to have a fantastic family life and all that's great and wonderful, but it wouldn't carry you through an entire year's worth of work. Because your work's so - you know - you're working from 8 to 5 or 7 or whatever it may be every single day. Unless you're happy in that I think it would be extremely difficult to just turn off and come home and be contented and happy thinking, "Oh God, I've got to go back there on Monday morning", and "Oh, God, another week of that". So that's the sort of reasons and I'm getting job satisfaction and I'm actually enjoying it. I'm going there and I'm not getting up on a Monday morning and thinking "Oh, no".

(Male / 46 years old / referred / IT project manager)

iv) Stress: The stimulation and stress of work was mentioned above but it is also worth recording this as a distinct sub-category when the emphasis was upon the deleterious impact of work-related stress. This may have been knee related, if interfering with work or simply due to other aspects of the work (e.g. returning to work following a career break). For some interviewees salient features of work relevant to quality of life were the quantity or structure of their work hours (for example, working full- rather than part-time or working shifts) or logistic considerations such as commuting.

v) Income generation: Work viewed essentially as a source of income enabling the individual to earn a living or a certain standard of living was included here. Work as a source of finance therefore was the starting point of other aspects of life:

I#31: I'll tell you what I like, I like my work, for a start, that's really important, my job because without the job I wouldn't have anything.

(Male / 57 years old / referred / newspaper editor)

and:

MR: Work was the first thing you mentioned in terms of your overall quality of life ... How does that affect your quality of life?

I#8: Money makes the world go around I think.

MR: Right.

I#8: Um, it's just you've got to work to get the nice things in life I think.

(Male / 39 years old / referred / chef)

The relation between work and money was so close for some people that one essentially was a proxy for the other:

II#2: Um, the world revolves around money which is work. The family's always there hopefully. And you need some 'social' and you need some 'health'.

(Male / 40 years old / listed / civil engineer)

vi) Social component: For some people, work provided an important social function either as a shared environment with friends or providing opportunities to meet new people through work.

II Family

i) Stability: Family concerns were mentioned as a quality of life domain by most respondents. Issues covered under this general heading included the importance of stability within family life and of there being routine. Stability and happiness of the family was viewed as the starting point for everything else and a buffer against external pressures:

II#22: Um, it's the biggest part of everything. Good family life, and everything else tends to flow from there. If you get in arguments or family life's going bad it kind of knocks on to everything else you do.

(Male / 34 years old / referred / carpenter)

Aside from health, general happiness within, and of, the family was considered important, and the negative consequences of problems within the family noted. Maintaining good relationships within the family was regarded as important to quality of life. The following two extracts illustrate some of these points:

I#11: I think as a family we are happiest when we've got a routine, we are not a chaotic-type family. We like things to be straightforward so we can plan ahead. I know it sounds a bit boring but we don't like unpredictability too much

#11: *Family - that means all aspects of the relationship between individual members of the family, their health and well-being obviously.*
(Male / 44 years old / referred / teacher)

ii) Immediate and extended family: The SEIQoL-DW protocol seeks to avoid individuals being provided as a cue. Nevertheless, this was occasionally the case. Furthermore, following probing, interviewees expanded upon who they were considering when referring to their family. Comments were made about both immediate and extended family members and included children, parent's siblings, grandchildren and parents, in-laws and cousins. One interviewee, a widow, described the enduring influence of her husband and his death (eight years previously) upon her quality of life:

#18: *I can't say I'm 100% happy because I lost my husband, you know so.*
(Female / 64 years old / referred / retired)

Partners formed an interesting sub-group and were perhaps noticeable by their relative rarity of mention. Occasionally relationship difficulties were discussed, for example, one woman describing the substantial negative consequences of a relationship split (#13). Another interviewee described the problems that a daughter with mental health problems was causing between herself and her husband (#39). Partners were also mentioned as sharers of a common interest or hobby.

Offspring, both child and adult offspring drew specific comment in relation to quality of life. Sons or daughters with emotional, mental and physical health problems were a source of concern (e.g. #13, #26). Broader concern and interest about educational, social and personal development for younger children was also described. More positively, children were also seen as providing an energising and uplifting boost. For older children, other issues emerged such as concern about their financial security.

Grandchildren played a similar role in quality of life terms as children, although possibly with a greater emphasis upon enjoyment. Some of the

interviewees did though mention a caring role in relation to their grandchildren. One interviewee described how he and his wife 'adopted' his nephew's children as grandchildren and realised the commitment that this entailed:

Wife of #5: ... we just don't really realise like all the implications of being a grandparent what it involves. I mean really it involves being second pair of parents to the baby really
(Male / 52 years old / listed / electrician)

The 'family' category was further sub-divided into the following headings:

iii) Family role: Interviewees were not always expansive about the way in which 'family' contributed to, or affected, their own quality of life, despite readily providing it as a cue. For others though, the nature of this relationship was made explicit and included several components. A supportive or caring function was identified with the interviewee either being the provider or recipient of support. The support could be practical in nature especially in relation to physical health problems, or emotional, including providing a sense of security (e.g. #1). One of the interviewees, a student, described his family's support as both financial and (specifically in relation to his knee problem) moral (#19). One extract summarised the importance of the family to one patient who had been badly affected by his knee problem:

#24: Um, the happiest part of my life is having a wonderful family.
(Male / 56 years old / listed / glazier)

Although perhaps reciprocal in nature (e.g. #32), such caring relationships were usually expressed as unidirectional. As a carer, the interviewee was most often referring to an adult family member, for example, sibling or elderly parent rather than a child. One interviewee who discussed caring for her children was in fact discussing her adult daughter (#35). Caring for an adult relative may be a responsibility that merits special relevance as a quality of life domain, whereas for a child is simply expected. For example, several such

accounts related to particularly challenging circumstances, for example, physical or mental health problems of a relative.

The practical consequences of caring for relatives were stated, such as regular visiting of elderly parents or (for example, #10) laundering clothes for an incontinent spouse. Such consequences were expressed both as burdens and as simple facts. In addition, a family member's health was a quality of life concern in its own right. Health of family members was expressed as a prerequisite or building block for happiness and well-being. This was expressed both in the absence of significant illness within the family, as well as its presence. In one case, family illness had been the stimulus to re-appraising life priorities:

I#17: ... we've had a lot of problems with my Dad - with health and the council and his aorta operation and that sort of thing. So it's brought a lot of things to the forefront in realising just how important things are in life rather than nit-picking over stupid things.
(Female / 51 years old / listed / school support worker)

The function of family members providing companionship or friendship was also evident. Partners and other relatives were described as close or even best friends and playing a large social role within the interviewee's life. The role of a partner within a sexual relationship was mentioned.

iv) Family activities: Occasionally the nature of family life as a quality of life component was expressed in terms of shared activities. This usually involved leisure or sporting activities with younger children and indeed, such shared experiences appeared to be the core of their domain:

I#2: Happiness within the family. My ability to be fully active and by that to be able to do some sort of physical exercise. That's all really just general social things you know like watching football and being out with my boys. The boys are my family.
(Male / 40 years old / listed / civil engineer)

and

Il#23: The most important really is doing things with the family - together.

(Female / 46 years old / listed / hospital cleaner)

v) Pets: were included under 'family' because of the role they played within the family unit. In some cases, pets acted as a facilitator or focus for family activity, for example, jointly walking the family dog. The necessity of walking dogs also provided, for some, welcome exercise and the pursuit of an enjoyed activity. Aside from such functional concerns, pets were simply a source of inherent pleasure as well as being viewed as a defining attribute of the family (i.e. this is a family that keeps pets). The significant time and care commitment was also a factor and the consequences for not maintaining these commitments was expressed.

III Money

i) Basic provision and debt: Money, and financial considerations more broadly, were mentioned by several participants. Issues raised related to what may be immediate considerations about basic provision and debt avoidance through to longer term considerations. The former concerns were raised by those currently experiencing or facing the prospect of protracted sick leave (due to their knee problem) without necessarily adequate sick pay cover. Money was raised as a consideration because of the worry it could generate and its negative impact upon a quality of life. Vicarious concerns were also expressed in relation to the financial difficulties of adult children. Experienced redundancy of a spouse was one stimulus for referring to finance.

ii) Enabling finance: Aside from these potentially negative aspects, money was important for its role in enabling everyday life functions, allowing mortgage payments, paying for everyday items and expenses. Generally these activities and functions were relatively modest and often interviewees stressed an absence of extravagance:

Il#12: Having some money. Not a lot of money, just having some. Enough to get by on so I can do what I want, when I want. Go shopping, holidays.

(Female / 29 years old / listed / customer relations officer)

However, perhaps naturally there were expressions of wanting (a little) more than they currently had. Several interviewees expressed the concepts of financial stability, security and planning, whether this was to accommodate recurring costs, or intermittent and episodic expenses. This included not just whether one was going to continue working and earning, but also that existing level of reimbursement was going to be retained. Finance enabled the freedom to conduct one's life with a degree of liberalism:

MR: Mm. Ok, you mentioned money. Again how does money effect thinking about your quality of life at the moment?

II#17: Mine? Without it I wouldn't be as happy as I am now. I like the idea of being financially sound, to be able to do what I like when I like. I'm not talking about buckets of money - like everybody else I'd love it.

(Female / 51 years old / listed / school support worker)

and:

MR: Okay, how would you describe what financial security means for you.

I#11: Well, just the fact that we feel comfortable enough to have the lifestyle we enjoy, not that we are rolling in it by any means obviously but our choices are limited by our income, the opportunity to travel that sort of thing, take the kids away, family holidays.

(Male / 44 years old / referred / teacher)

iii) Long-term security: A slightly longer term financial perspective was indicated by some interviewees who referred to planning for retirement. One example:

II#9: Well as I'm getting older, what I'm trying to do is build myself a firewall of income.

(Male / 51 years old / listed / self-employed businessman)

The same interviewee, a businessman with a clear entrepreneurial instinct described a particular interest in finance and familiarity with the broader financial world. Money was not simply a means to an end, a way of coping

with everyday day expenses – he had an interest in the mechanisms of the financial world:

II#9: Much long term, in as much as I've got a portfolio of stocks and shares, I've got a small portfolio of properties and we're looking abroad to buy something both to use and to rent.

(Male / 51 years old / listed / self-employed businessman)

His comments about his financial status within his retirement also reflected the relative perspective that individuals held about their own situation:

II#9: So – yes, that's what my firewall – I've actually basically done my own thing and I've got my own streams of income which, you know, are not great. I mean it wouldn't give me a good standard of living but it would keep the wolf from the door.

(Male / 51 years old / listed / self-employed businessman)

IV Social life and friends

Several concepts have been included under this general header, including 'friends', social activities and functions. Some of these areas appear to show a degree of overlap with quality of life areas categorised under other headings – such as sporting activities. However, for such cues under the present header, the emphasis is upon the social component of the activity rather than, for example, the intrinsic sporting element. Defining 'social life' is complex and was a question posed to interviewees when providing the cue label. A rather broad and exclusive definition was provided by the following interviewee:

I#2: Anything that's not to do with work - going for a walk up the mountain; driving somewhere; driving somewhere for the weekend with the family; going to a football match - anything other than work.

(Male / 40 years old / listed / civil engineer)

and a narrower one:

MR: Again what's social life for you, if you like?

II#22: I've got a top pub 100 metres away!

(Male / 34 years old / referred / carpenter)

i) Social activities: The variety of activities included under this header were dining out, or with, friends, structured entertainment such as cinema, clubbing and going to pubs. The latter could be regular mid-week visits to the local pub for a quiet drink, an unwinding at the end of the working week, through to intense partying. The variety and contrast within even the same nominal activity is illustrated by the two following quotes about going to the pub:

I#33: Um I don't know. Not really, I just like to, you know obviously I like to enjoy myself as well so... you know partying hard.

MR: Ok so social - .

I#33: Social life yeah, does mean a lot to me so.

MR: So what does 'partying hard' mean for you?

I#33: Well there's, you know there's the lads on a Saturday night. I've always gone out once - I play skittles – once, twice a week I've always got to go out and meet up with the same sort of rugby mentality really, sort of lads so.

(Male / 38 years old / referred / builder)

and:

MR: Now you also mentioned your social life which you said is non-existent.

I#10: As I say, all I do is sometimes I see my friend on a Tuesday and a Friday for a couple of hours and I go out for a pint or two and that's it.

I#10: Yeah, just sit there and have a little chat about football and what's happening on TV, that's all.

MR: Yeah.

I#10: Just you know, that's all.

(Male / 51 years old / referred / retail worker)

ii) Level of organisation: Social activities could be more or less formal in their structure and organisation. Formal social activities mentioned included Tangent meetings (for wives of retired members of the Round Table) and also a church youth organisation (Navigators). Less structured activities included those described above and also some other less familiar 'social' activities such as shopping. The following extract illustrates this aspect for one interviewee who cared for her adult brother and for whom her knee problem (and other mobility difficulties) was having significant negative impact:

I#21: *I know it sounds silly but I used to enjoy going to the shops –.*
MR: *Doesn't sound silly at all (laughs).*
I#21: *I used to love going to the shops and getting on the bus and going out visiting and just generally doing for my brother. I used to take him out.*
MR: *So when you say you used to enjoy going to the shops, do you mean ordinary grocery shopping or just –?*
I#21: *Oh I used to like going into Cardiff - window shopping and things like that, but I don't do it any more.*
(Female / 49 years old / listed / housewife)

iii) Following sport. *Playing sport* was included under a separate quality of life header but *following sport* was included here. There were three discernable aspects. Firstly 'remote' viewing involved watching sport on the television (e.g. at home or in company, at a pub). Secondly 'near' viewing included attending sporting events either as a casual observer or more consistent supporter. The third characterisation was 'general' supporter which indicated an overall support for a team but didn't imply how that support was pursued. Of course, within each of these aspects, there was more involved than just watching sport and may have implied other activities such as drinking and being with friends. Example extracts illustrate the second two aspects of following sport:

Wife: *I tell you what I'm surprised that you haven't put on there is um, your alcohol in fairness.*
...
I#36: *Well that comes under sport [laughs].*
MR: *What does that mean then?*
Wife: *Well he likes to go out with the boys once a week at least which he hasn't been able to really - which is very important to him.*
I#36: *Yeah, I miss a bit of that but, that's the same as football, that goes hand in hand with football doesn't it.*
(Male / 40 years old / listed / builder)

and also:

I#32: *I sometimes get to go [to watch rugby], but more often will get together with my brothers and my father and go down one of their houses and kick [the] women out for an afternoon. Make them grumble for a while. Magic!*
(Male / 46 years old / referred / IT project manager)

iv) Friends: Although social life was synonymous with friends for some individuals, they were often discussed as separate but related concepts. Friends could be defined as a close circle of long-standing friends, as well as a larger circle of contemporary acquaintances. For some, friends may be old school friends; an amalgam of prior and current work colleagues; neighbours; or definable within clearly distinguishable social groupings:

/#19: Most probably separate. I kind of tend to kind of have separate groups, like medic friends who I see on an individual basis, then church friends and Navigator friends and then kind of other friends that I live with and people like that ... my friend groups don't overlap much, they are very separate.

(Male / 24 years old / listed / medical student)

Friends provided a number of functions relevant to quality of life. They were a source of entertainment, companions within shared activities and people one may simply pass the time with.

/#15: Well you get support from different things from family and from friends. Obviously friends provide a bit of entertainment as well whereas family less so.

(Female / 19 years old / referred / medical student)

For some of the younger people interviewed, friends served as a distraction from the recent separation from family - even providing a surrogate family. Friends were a particular consideration for one interviewee who had moved (from his parental) home in the preceding year:

/#16: Well family and friends are probably the most important thing. Keeping up with them. I moved here about a year ago so I've seemed to have made quite a few friends since being down here so that's made me pretty happy, so.

(Male / 28 years old / referred / administrator)

Friends provided support; a peer group of people with similar challenges; and interests with whom one shared ideas. They helped to share burdens and

confidences. This was especially important if the burdens shared were difficult to share within the family:

II#17: Yes it's nice to have friends, be able to talk to different people. Share - share things with. Sometimes you can talk to a female friend when you can't talk to your husband. Nothing personal but - just silly little things, that sort of thing. You don't want always to talk to your own family either so it's nice to be able to have friends.

...

II#17: Cos you don't want to talk about it within the family because you don't want to upset the immediate family by sharing some of your fears then shall we say.

(Female / 51 years old / listed / school support worker)

V Health

As study participants were being interviewed because they had an on-going knee problem, health was likely to be a key quality of life area for at least some people. Whilst the stimulus for forwarding health as a cue was the present knee problem for some, other concurrent medical problems were also a reason for discussing health. Mostly interviewees discussed health in relation to themselves but health of a family member was also mentioned as effecting quality of life. Having good health was seen by some as a fundamental pre-requisite for good quality of life:

I#30: If you haven't got your health, it affects all areas of your life. It's no good having everything else in life if you haven't got that ability to benefit from it.

(Female / 44 years old / listed / residential care team leader)

and:

II#15: Well if you're not healthy then you haven't got much quality of life really have you in a sense. If you're ill then you can't do much else.

(Female / 19 years old / referred / medical student)

and finally:

I#5: And I think the two go together. If you're healthy you're happy and vice versa.

(Female / 19 years old / referred / medical student)

i) Enabling health: Health as an enabling and liberating state was described by several interviewees – being fit enough to get on with life's activities. Good health allowed freedom, independence and the ability to do things without restriction or worry. Poor health meant restriction, increased reliance and a detrimental impact upon mood. Concern raised by a specific health problem may not even be related to its current impact but an awareness of how ill-health in general may disrupt the future. Being sufficiently healthy to care for other family members was expressed by some interviewees (mainly mothers). This was particularly key for one single mother. Poor health was seen as a restriction upon this role and therefore damaging to family relationships and overall family well-being. One male interviewee who described this aspect included the consequences for work and social life within his consideration of family life. The following two extracts illustrate some of these points:

I#1: Well, I think more freedom, I would be able to go out on my own and not have to rely on other people to be there with me.

MR: And that's related to your health.

I#1: Yeah, well health and other reasons but basically health, you know the problem with the knee at the moment with it giving out. I'd be walking along the road one minute and I could be on the floor the next, so and unless I've actually got - having to rely on someone – it would be nice not to rely on people all the time.

(Female / 41 years old / listed / retail manager)

and:

MR: Yes okay, so your health is one very important component of your life at the moment?

I#35: Only because without my good health I can't care for my family.

(Female / 63 years old / referred / retired social worker)

ii) Physical, mental and functional health: Health was represented slightly differently by those discussing it during the interview. One characterisation of 'health' included for one interviewee both leisure and sport. More commonly

though, interviewees discussed physical health; its functional correlates; and mental health issues. The latter included effects consequent to their knee problem (for example, difficulties with pain affecting sleep, or anxiety about returning to previous levels of physical functioning). For one interviewee, health incorporated a novel sense of general vulnerability resulting from their knee problem. Other physical health problems were also mentioned. Attributes of mental well-being not linked to physical problems were also discussed. One interviewee described the importance of protecting time for herself during the week. This was an attempt to remove herself from the daily and family pressures in her life and provide an opportunity to relax:

I#30: Yes, time to myself with the exercise. People might say that if you can get to the gym 3 times a week you're lucky, you get time for yourself, but to me that's a necessity.

(Female / 44 years old / listed / residential care team leader)

Health as physical well-being was mentioned by several interviewees, either in relation to the knee or other health problems. Functional problems were highlighted including general mobility and restrictions in getting on with everyday life. Weight gain secondary to functional limitations was an issue. Excessive delays for intervention were also mentioned in this regard, as were concerns about the need to exclude malignant pathology in relation to an initially undiagnosed knee problem. The presence of significant co-morbidity was noted by some as a major source of quality of life concern. Knee-related mobility problems also brought in perceptions of aging as a health located issue, as well as the importance and role of independence. The following three extracts below illustrate some of these points (perceptions of aging and independence are explored more fully elsewhere):

I#38: Generally I can get on with things and I've always been healthy and you know got on with life. The knee at the moment, well both knees actually are causing a problem 'cause they restrict what I can do.

(Female / 47 years old / referred / teacher)

and:

MR: *And so health again that was the fourth area that you mentioned.*

I#5: *It takes I'd say 70% of my life ... I mean I've had about eleven operations throughout. [operations included some on back and other clinical sites]*

(Male / 52 years old / listed / electrician)

and finally:

I#5: *I'm very disappointed with my standard of life regarding the health side of it and I feel like I'm ageing faster than I should due to the lack of mobility.*

MR: *Would you say that the feeling of ageing... that's really much a part of your health?*

I#5: *Very much.*

MR: *Yeah, it's not something that you consider as a separate area of your life?*

I#5: *No I think it's the health side of it. Yeah, definitely affects my standard of life and my enjoyment or quality of life I suppose would be the right word.*

(Male / 52 years old / listed / electrician)

iii) Negative and positive health: Prior experience of ill-health, perhaps unrelated to the knee problem was influential in health being mentioned as a cue. This included illness suffered by the interviewee or their experience of friends and family with, for example, terminal illness. Therefore, health encompassed a desire to both avoid ill-health, as well as to return to normal levels of function and health. A more positively framed perspective included achievement and maintenance of (for example, cardiovascular) fitness as opposed to simply the absence of disease.

VI Physical activities

Whilst there was some crossover between content areas included under this header and others (e.g. health), the fundamental nature of what was described marked it out for separate consideration. Under this header, further distinction was made between *sporting activities*, *fitness activities* and *general physical pursuits*.

i) Sporting activities: These included both team and individual sports with varying levels of formality; activity level; and competitive structure. These included familiar sports such as football, rugby and snooker and less common ones such as dragon boat racing and sea fishing. Some activities such as competitive running were described by at least one interviewee as 'sports', although running for others may have been regarded as a fitness pursuit.

Sport provided many functions including competition and companionship. Competition may have been at a relatively high level (e.g. county sports) or simply with friends and colleagues. For many it was an opportunity to make and spend time with friends. Allied to this was the importance to some of participating within a team and the value of a shared team spirit. Whilst a sport may serve one or other purpose more strongly for some interviewees, for others the value of sport was multifaceted:

MR: ... so in terms of what you get out of playing sport, what do you think that, you know, what are the benefits to you participating in sport because I think there are a variety of reasons people play sport?

I#15: Yes its got to be a team sport, I mean its just the team spirit of the game I think.

MR: Right.

I#15: And winning obviously [laughs].

(Female / 19 years old / referred / medical student)

and:

I#19: Just something that friends do, it's just the team kind of aspect, fitness as well kind of being able to do something that keeps me refreshed and its good for stress as well because I used to run about three times a week on my own, not like competitively or anything just for my own kind of fitness and to de-stress myself on.

(Male / 24 years old / listed / medical student)

Both interviewees quoted above were students and described the mental benefits of sport, in addition to the physical consequences. The second student expanded on this by saying how sport helped him feel less tired, think quicker and remain alert. He described how sport benefited his self-esteem. Others described how sport helped with work or family stress. Sport though

was also a means of expanding family life, for example, playing sport with children. One man described the freedom to contemplate and resolve problems whilst out marathon training. When the ability to undertake sport was either removed or compromised its loss was keenly felt:

MR: ... how important is sport for you in combating stress of everyday life if you like, stress of your studies?

I#15: Very important because you've got to go out and burn off energy - because it just burns off the energy excess.

MR: Mm.

I#15: Energy, sitting at a desk and revising or whatever, you've got to go and do something then to take your mind off it.

MR: Right, has that always been the case for you, sport allows you to do that?

I#15: Yeah.

MR: Ok, so when you're not playing sport, how does that affect you?

I#15: Aggravating.

(Female / 19 years old / referred / medical student)

Sport was something that some people had grown up with since childhood and was a familiar and integral part of their life. One interviewee, a former soldier, described how his sport was a continuation of his active and sporting life from his service time. Sport, intentionally or otherwise provided structure to people's lives. The pivotal nature of sport within some people's lives was reflected in the following comment:

I#33: That's right yeah. Well 'cos most of my life was built around sport and I, you know the rugby club and everything like that and that's my focus.

(Male / 38 years old / referred / builder)

ii) Fitness activities: This theme included going to the gym, running, skiing, swimming and walking a pet dog. The absence of a competitive element marked these activities out from those described above. In contrast with comments above about sporting companionship, for one person at least, going to the gym was an opportunity to be alone. Nevertheless, some common functions were served including being an aid to stress management, a shared family activity (or a means to escape from family pressures) and a means of simply feeling good. Similarly, benefits to overall health and fitness

were mentioned. In addition though, weight reduction and management was more likely to be discussed within this category. This was in some cases directly related to the presenting knee problem. Exercising provided a sense of achievement – especially if attempted weight reduction was achieved. The extract below illustrates the place of physical exercise in one interviewee's life:

I#30: Since I've lost weight and that's helped my knee, I've been fanatical about the gym.
(Female / 44 years old / listed / residential care team leader)

iii) General physical activities: Here the emphasis was on being physically active rather than any particular form of exercise or sport. This was important for a few interviewees who had already experienced mobility problems due to their knee or other health problems. This focus on being physically active was expressed through a variety of outlets such as walking, gardening or dancing. For one woman (a widow) cycling was a major part of her life. It had been a shared activity with her husband and a means of holidaying when he was alive. Pragmatically, it now remained as means of transport, and symbolically as an expression her identity and independence.

VII Leisure and Pastimes

Within the overall thematic analysis of quality of life, sporting and social activities have been included under their own separate heading. Other activities largely pursued for pleasure have been coded under the general heading of leisure and pastimes.

i) Gardening: The presence of gardening as a quality of life cue represented several factors. Strong expressions of attachment to gardening were evident from interviewees whose mobility problems restricted this activity. One woman described how she loved gardening but had paved over some of her garden due to the increased challenge of maintaining it adequately. The practical necessity and responsibility of attending to a garden was also discussed and

in this regard it may be distinguished from other activities described under 'leisure and pastimes'. However, the burden of gardening was introduced in part by the development of a mobility problem and other activities also entailed a notion of commitment and obligation.

Beyond this though, gardening was described as an activity shared with a partner; a means of distraction; relaxation and escape; as a source of satisfaction and achievement when projects literally come to fruition. Although one interviewee mentioned the pleasure of visiting show gardens this was distinguished here from gardening as a process. Gardening as an activity, of course, fluctuated with the seasons and its significance varied accordingly.

The following extract illustrates some of these points:

MR: Mm, okay. One thing you mentioned was gardening - was something that you really enjoy doing.

I#39: Mm I love gardening.

MR: So is that something that you might regard as being something that's important to your quality of life?

I#39: I think so because it's about the only thing I do that I lose myself in.

MR: Right.

I#39: Yeah, totally.

MR: What do you mean by lose yourself?

I#39: You just completely de-stress and it's just lovely, I love it.

MR: Mm.

I#39: Really, really love it. I'm not very good at it!

(Female / 50 years old / referred / administrator)

ii) Hobbies: A number of hobbies were also described, including making tapestries and collecting, the latter being mentioned by two interviewees. One of these people described how collecting feline figurines provided a number of functions in her life, in addition to an existing interest in cats. It was a shared hobby with her partner and sister-in-law; it provided a stimulus and framework for social outings, it gave her intellectual stimulation; it was a distraction from pain and she admitted that it was also addictive. In the face of an otherwise diminished social life she commented:

I#1: Yes, that's one of the enjoyments I do have left.

and:

I#1: And sort of then throughout the week I'm on the internet and I'll check up on the cats that I bought and whatever and it gives me something to look forward to then on the weekend.

(Female / 41 years old / listed / retail manager)

Hobbies provided a sense of achievement, noticeably so in those that involved producing something tangible. One man described the pleasure he derived through renovating vintage commercial vehicles. He also valued the ability to immerse himself without distraction in this work, which also saw him regularly travelling to various shows. The woman above whose hobby was making tapestries revealed her satisfaction from the product of her activity:

MR: And how long have you been doing your tapestry for?

I#18: Oh God, years, years. Yes, years. My husband bought me that one – 'The Last Supper'. So I managed to get that finished. And I took it to where it had to be framed, 'cause my husband used to frame them for me, and I thought it's going to cost me a fortune. I went down and I said 'How much will it cost me to have it framed?' and she said, '£45', and I said 'Do it!' I was quite a way from here and when they showed it to me after it had been framed, I was so pleased with it, you know and he wrapped it up and he said 'Where do you live?' and I said '[name of residential area]'. He said would I like him to drive it – 'No', I said "I'm going to carry it all the way home". I was so pleased with it, you know [laughs].

(Female / 64 years old / referred / retired)

iii) General leisure: A variety of other leisure pursuits were mentioned including driving, computer activities (such as internet surfing and gaming) and music (both listening and singing). These and other activities such as dancing and day trips with the family were included under this sub-header rather than elsewhere because their focus was on leisure rather than any other role or function. For example, one interviewee described his reason for nominating listening to music:

MR: OK and is there a final thing that you can think of that we can put down?

I#7: I don't know ... well I like my music – bit of jazz, bit of blues but that's just pure leisure that is.

(Male / 53 years old / referred / local government administrator)

VIII Personal environment

This category represented aspects of the interviewee's immediate or broader environment, and included physical, social and psychological factors. Most of the content related to the immediate home environment.

i) Home ownership: The first theme was conceptual rather than practical. One man described the importance of his house as an investment and the associated necessity of maintaining it adequately. A second interviewee described the importance to her of actually owning a property. She explained it was significant because it was her first house, and bought jointly with her partner. She further described the importance of owning something other than a car:

I#12: You know, I get too stressed out with that, but just pitching around and knowing that we actually own something other than a car is nice.

MR: So the fact that you feel, or the fact and the feeling you own it yeah?

I#12: Yeah, it's nice to just come home and you can just relax and not have to worry about it.

(Female / 29 years old / listed / customer relations officer)

ii) Emotional appraisal of home: A second theme related to an *emotional appraisal* of the home, for example, as a place of sanctuary. The home was a place to enjoy, to retreat to, for example, away from work and provided peace and relaxation. It was also the location for the conduct of activities such as cooking which may be valued as an enjoyable pursuit or, for example, as a role within the family (as mother, wife, cook etc). For one interviewee who had recently moved, 'being settled' was also an important aspect of the home. The home was seen as a place to be with the family or a partner and may have been used as a proxy for expressing the importance of that relationship. Emotional attachments to particular properties were strong for people, for example because of its role as a family home or duration of occupancy:

I#35: Well to me that's important, to me to give up this.

MR: Yes.

I#35: *Just the cost of the house. Yes I could afford to sell this house and buy something quite suitable for myself but I don't want to, this is the home that my children come back to, "lets go to nanny's and bompys", that's it.*

(Female / 63 years old / referred / retired social worker)

The home offered safety particularly for some interviewees experiencing mobility problems or who felt restricted to their homes. In contrast, features of the home itself also presented mobility challenges for the interviewee. For one person who had suffered from numerous physical health problems, being away from the home for an extended period of time presented a real challenge:

I#5: *For a start, I like to cook a lot. So I think I'm generally happier at home than anywhere else.*

MR: *Right.*

I#5: *I mean when I'm on holiday I get homesick to come back.*

MR: *Right.*

I#5: *And I think that's part as well as a health side. When I'm away I start to panic. 'What would happen if my back goes or if my knees go?' Not so much my knees but my back especially.*

(Male / 52 years old / listed / electrician)

iii) Home maintenance: Owning a home brought domestic responsibilities which were relevant to quality of life. It had to be maintained out of necessity, interest or choice. This included both housekeeping and decorating. For some people, having a clean and tidy house was an important consideration and maintaining it part of their role. Being unable to contribute towards this due to health problems was consequently a source of frustration. The necessity of decorating following a move to the owner's current house was described:

MR: *So you've obviously done quite a lot of work on your house?*

I#5: *When we bought it, I mean it was exactly as they bought it – there was no central heating, they hadn't decorated - they painted around wardrobes and bed and things which I never thought, I never thought really happened but the classic was the bedrooms. In order for the doors to - instead of shaving the bottom of the doors to get it move over the carpet they actually did cut semi-circles out in the carpet. Honestly, I'm not lying! I'm not telling you stories. Honestly they actually did!*

(Male / 52 years old / listed / electrician)

iv) Neighbourhood: The broader environment was also discussed, and included an expression of disaffection with undesirable social features of the local neighbourhood by one woman. She was particularly concerned about the potential impact of local social problems upon her son. More positively, two other interviewees described their satisfaction with the local environment. For example, the first was happy with access to and from his home to important destinations. The second, who was temporarily living away from her parental home, drew satisfaction from the social and physical environment that she was currently inhabiting, and the lifestyle opportunities it afforded her.

IX Psychological constructs

Throughout other categories of the overall framework various abstract and psychological constructs were described, often as consequences of, or underlying, described activities. However, in some instances these states were addressed directly by the interviewee and were included here. Some of these were idiosyncratic and may have been described by only one interviewee. The concepts also varied in their degree of structure when verbalised. On occasion, somewhat diffuse descriptions were provided by interviewees. When this was the case, the description was characterised and labelled during the interview and passed back to the interviewee for their endorsement.

i) Independence: This was mentioned by a small number of people as a quality of life area. It was also a construct that was referred to by several others during the course of the interview. Independence involved the freedom to do things oneself and not to be reliant upon others. This included self-care (for example, bathing), as well as other family and domestic activities. This situation was described by the following woman:

II#1: And I can't stand long enough to do it. So it would be nice to sort of have the freedom of doing things that I used to do before.

MR: Yeah.

II#1: Like I said before, playing with my grandson. I can't kick a ball with him. If I get down on the floor to play with him, after about 5 minutes I got to get up because its so uncomfortable.

MR: Mm.

II#1: And it's just taken the pleasure away.

(Female / 41 years old / listed / retail manager)

The ability to function without support may have been elicited as a cue due to an interviewee's experience of independence being compromised (by illness or other cause). After all, it wasn't mentioned by others who may well have considered their independence intact. A second interviewee described her desire to 'to have a life' – to live without the current substantial restrictions that she was faced with due to a catastrophic relationship breakdown.

Encompassed within this definition was her need to retain her self-identity and to be independent of other influences, a second and different aspect of independence than described earlier. The following two quotes from her reflect this and also her endorsement of the independence label:

II#13: Not to lose the identity of me as a being - which is unfortunately what's happening with everything else.

and:

II#13: I would say having a life would be independence.

(Female / 32 years old / referred / retail customer care worker)

The described consequences of losing one's independence were subtle making it easy to overlook its significance to an individual. One woman whose mobility was severely restricted and who was largely dependent upon her husband due to substantial co-morbidity described her inability to fulfil her family role in grocery shopping. The significance and pleasure of the 'little treats' she described may not be adequately represented by simply noting their absence or superficial function:

II#28: And whereas if I went shopping I could go around Sainsbury's at my pace and you could look - and now you have a list and stick firmly to that list and nothing else goes into the trolley. Whereas when I used to do it by myself a little treat would go in for us cos just the two of us at

home, but little treats don't go in no more. Probably, my husband hates shopping so the quicker he can get around and out the better.

(Female / 61 years old / listed / former nurse)

ii) Achievement: A second construct was *goal-setting* and achievement, described by one man. This involved the explicit laying down of written goals on a biannual basis. The interviewee noted the 'buzz' of achieving such goals and his orientation to any failure to meet goals which was then regarded as a 'learning experience'.

iii) Adventure: One young woman described her desire for exploration and, as she termed it, *adventure*, which manifested itself in travel. She discussed how she could be dissatisfied with familiarity, of being in the same place for too long:

MR: *Give me an example of what you mean by 'exploring'.*

I#6: *OK, the one example is travelling – I did last year or the year before actually got a map and mark everywhere in the world which I wanted to go and I do plan on going everywhere at some point. I don't know, I just like doing different things and meeting different people and I wouldn't like to be stuck with the same old boring people and same things – I don't think I could really have a normal office job where I'm stuck doing that, I mean I'm doing something like that now but at least there's a bit of variety when you get people screaming at you with it.*

(Female / 19 years old / referred / call centre operator)

Further examples of constructs that were alluded to within other categories were **iv) relaxation** and **v) stability**. The former was described by an interviewee and was achieved through activities such as watching the television. The latter was presented as an all-embracing construct which ran through various aspects of the interviewee's life:

MR: *The third thing you mentioned then was stability, can you tell me what you mean, you know what - .*

I#17: *Just stability generally, you know, work, home, you know like lots of my friends are divorced and this sort of thing and single sort of relationships and its just to me, I think stability is quite important. Especially I've got a 10 year old as well.*

(Female / 51 years old / listed / school support worker)

Another interviewee nominated *vi) social harmony and peace* as an important quality of life cue. This was in part her reaction to current and recent war and unrest around the world, and a concern for the state of the world that her descendants would inherit. Finally, in this section, someone described *vii) 'getting on with life'* as a cue. Although somewhat amorphous when elicited, the interviewee explained how this meant conducting her life and satisfying various quality of life goals without major disruption due to, for example, illness.

X Education

A small number of interviewees described education as an important quality of life area and were either students themselves or described the importance of education for their own children. For the students, their education was significant on a practical level because it was time-consuming.

i) Establishing a career. Perhaps more importantly, education was a means of improving their own prospects and *establishing a career*.

I#15: So I think it's basically because it takes up a lot of my time and that's mainly my future I suppose so.

(Female / 19 years old / referred / medical student)

Such a career may have been a long-standing ambition. In one sense education was synonymous with career and its nomination may have reflected the life-stage of the interviewees. This is suggested in the following extract in which a student described his current view of education having already completed a first degree:

I#19: I think education comes last for me because - partly because I have done a degree already and this is a second thing, so maybe at the back of my mind I have always gone along thinking if - I don't know - I don't get through, or get thrown out or whatever, I have got something to fall back on so it's not something - I mean it is important to me but it's not something that is overly important because I have already kind

of got my education. Maybe if I was doing it for the first time through it would have been a different ball game.

(Male / 24 years old / listed / medical student)

Educational achievement, and associated stress and strains of a demanding course (for example, medicine) affected quality of life. Another student (aged 24 years) described how education was important for her because she felt she had missed out going to college when younger due to protracted knee problems.

ii) Education for life: One mother and music teacher described the value and importance to her of education for her children. Within education she included not only school and university, but also a broader sense of her children being brought up 'properly'. That her children were educated in music was encompassed within this definition. One student also described spiritual education which he pursued by membership of a church youth organisation called Navigators. This provided the opportunity for bible study and a consideration of life and personal issues:

#19: Yes, just once a week, I mean that is just kind of a worship side of it, whereas the Navigators is more of a kind of learning side.

(Male / 24 years old / listed / medical student)

XI Spiritual and religious

i) Personal faith: A number of interviewees described the importance of spirituality to their lives and aspects of the institutionalised expression of faith within a religion. The importance of a *personal faith* was described and provided a base for approaching life and as a source of strength. One interviewee reflected upon how his faith provided a framework for approaching life and helped him deal with various difficulties. He noticed that when he had any problems with his own faith, he didn't deal well with things generally. He felt that if he had a problem with his religion it could impact upon his health and vice versa. A second interviewee described how her faith helped her:

I#38: Well it's all bunched in together but we've got a personal faith as well which means that if things do go wrong then the first thing we do is not panic but pray and I think that does give you a different perspective in life.

and:

I#38: Quality of life for me is quite a difficult concept because I've never thought about it in a concrete way because I've never, I don't feel I've suffered at all you know and so it's all come... you just grow up and get on with life and you just go through things. The Christian aspect has coloured everything we've done so... and in a way it does affect our quality of life because we've got a base and it sort of gives us a strength.

(Female / 47 years old / referred / teacher)

ii) Religious commitment: Having a faith and following a religious path was a long-standing *commitment* and time was considered an important aspect this. Religious activity was a significant part of the week at least in terms of the time devoted towards church attendance and other activities. Sometimes balancing all of one's commitments, including church attendance was an issue. Other activities associated with a faith could also be time-consuming and equally significant in terms of life quality. The Navigators organisation described earlier had a spiritual educational role, but also a social role within which the interviewee had a formal responsibility:

I#19: And Navigators just because its always been - since I've been at university really - its been a big part of my life.

(Male / 24 years old / listed / medical student)

iii) The Church community: One interviewee, who was experiencing substantial mobility problems, described the importance of attending church services. However, she was unable to do this because of the discomfort of sitting in church pews. She had settled for watching church services on the television. She also expressed disaffection with the church vicar who she felt had failed to fulfil a commitment to visit her at home. She also felt similarly let down by other members of the church congregation. She described herself as

having opted out and being disinclined to attend church, where she feared she might be the unwilling recipient of (excess) attention.

The formal church therefore also represented a *community*. One interviewee described the importance of Christianity to the whole family. She detailed their family involvement in church life; how the church was a social resource; and the importance of friends within their religious community.

XII Holidays

i) Family and social functions: The importance of holidays to quality of life was described by several interviewees and encompassed a number of distinct dimensions. Holidays permitted certain social and family functions - providing opportunities for protected time with the family or to make new friends and re-acquaint with old friends. Holidays for one interviewee, who originated from abroad, were synonymous with visiting family members. The value of holidays was especially evident for one family where the interviewee worked particularly long hours:

II#36: So holidays are really important as I say, because we do have quite a few don't we?

Wife: Well we work so hard as a family it's our only release and the only time we get as a family together, to be honest. Because my husband works 7 days a week, you know, so - so we did take the risk and do it [book an expensive family holiday] so after the holiday it will be constant cramming to save money for when he's off for the six months.

(Male / 40 years old / listed / builder)

ii) Escape and distraction: Going on holiday provided a distraction from everyday, immediate concerns and the domestic environment. Having a holiday to look forward to provided a means of dealing with the concerns of the day. This sense of 'getting away from it all' included the deliberate aim to engage in one's holiday to forget about daily life. The following two extracts illustrate these points, the former patient referred to the process as escapism and the latter described the function holidays have in removing themselves from their home environment:

II#35: Escapism. It's just my husband and I and I'm away from the external pressures like my daughter saying, mum can you pick [name of grandson] up from school.

(Female / 63 years old / referred / retired social worker)

and:

II#29: Yes, because we live on site, we're surrounded by these children all day and every day. Once they go home, we get an hour and a half's peace then we've got a youth club opposite which is a pain and then if we're on site, teachers expect to come in on a weekend[wife of school caretaker]

(Female / 54 years old / listed / catering worker)

iii) Mental and physical health benefits: The warmer and sunnier climate of some foreign destinations enabled relaxation and eased certain health conditions including arthritis and knee problems.

iv) Tourism: Some interviewees described a more positive perspective on holidays - valuing the opportunity to sight-see; experience different cultures; engage in physical and sporting activities; and explore. However modestly engaged, holidays were a luxury for some. The following two extracts illustrate some of these points:

II#6: Just um , yes there's so much stuff to see that its just depressing not being able to see it, there's so many amazing places it would be wrong not to see them I think.

(Female / 19 years old / referred / call centre operator)

and:

I#29: That is our bit of luxury you see.

(Female / 54 years old / listed / catering worker)

7.4 Discussion

Overview: Twelve major quality of life themes have been described for a sample of patients with mechanical knee problems following the application of SEIQoL-DW with associated in-depth probing of responses. The thematic representation reflects the analyst's interpretation of the cues provided by respondents, rather than their own particular labelling and categorisation. Nevertheless respondents' own descriptions were key to thematic formulation. The scope of the framework is broad and includes more than simply health-related elements, reflecting the individualised approach to concept elicitation. It therefore presents a rounded picture of what contributes to self-defined quality of life for this patient sample. Nevertheless it shows how knee problems have impacted upon patient quality of life both directly and indirectly. Furthermore, it describes the complexity and multi-dimensionality of the quality of life construct, the interrelationship between domain areas and its inherent dynamism.

Quality of life assessment in patients with knee problems

How do the identified areas of quality of life compare to those applied in the assessment of knee patients? Outcome assessment for patients with mechanical knee problems has traditionally focused upon mobility and physical functioning. One recent review of 54 outcome measures for anterior cruciate ligament deficient knee identified only two that were adequately validated: the Lysholm and Tegner activity scores.²³⁹ Both measures primarily address knee function - the former, activities of daily living and mobility - the latter work and sports related activities.^{176 251} In a subsequent review, Irrgang provided a definition of health-related quality of life as an individual's perception of their health, which again focused largely upon disability.²⁵² Although Irrgang acknowledged the multidimensional nature of HRQoL and that it attends to negative and positive aspects of life, the knee scale described only addresses change in symptoms, function and sports activity.

Broader influences upon life quality: However, there has been an increasing interest in incorporating complementary patient perspectives alongside traditional measures in assessing the outcome of orthopaedic surgery.²⁵³ Individual variation in expectations and outcomes for patients with different conditions may result in differential use of outcome measures according to whether one is assessing athletic patients and those with degenerative conditions. Garratt and colleagues reviewed 16 measures of health and quality of life designed for completion by patients with knee problems.¹⁷⁴ They noted that most measures addressed mobility, physical activity and activities of daily living. However, some measures such as ACL-QOL, EKFS and KOOS, assessed broader concerns (such as lifestyle, role limitations, social and emotional functioning).^{177 254 255} Whilst it is clearly appropriate that evaluation should focus directly upon the most likely consequences of an intervention (e.g. mobility), the current data serves firstly as a reminder that such functional aspects are manifested in diverse ways and may exert variable impact upon individual quality of life. Secondly, patients should be appraised in a rounded context, which may or may not include consideration of their health and its consequences.

Knee-related quality of life: Within the current analysis, determining the consequences of the patient's knee problem for their quality of life was not an objective, although it may have been suggested by the patient in the interview. Thus descriptions in the results above show how unresolved knee problems had often profound effects for patient physical health (e.g. restricted mobility, weight gain) and mental health (e.g. depression, perceptions of aging), as well as other areas of functioning such as family and social life. The patient's narrative account of their knee problem and its treatment formed the latter part of the interview, and patients were directly probed about the broad impact upon their well-being. Whilst this data will be reportedly separately, it falls outside of the remit of this thesis.

The broad content of the thematic framework and the areas of life represented are not unique and many have been reported in other studies using the SEIQoL approach. For example, Campbell and Whyte also reported domains

such as family, health, work and independence in their study of cancer patients.²³⁶ Interestingly, they also noted the nomination of individuals within the family as cues; the differing composition of 'family'; and that different elements of the family may be provided as separate cues. In a study of patients undergoing hip arthroplasty, Bayle reported 70% of respondents nominating health, 90% family and 30% work.²⁵⁶ However, as is clear from the current thematic framework, comparison across, and even within studies, based upon such main category headers may be misleading. For example, the content of health concerns expressed by patient in the three studies may be quite different.

The nature of described quality of life content

Causal or indicator variables: The classification of emerging quality of life themes raises questions about the nature of the cues described by respondents. The first relates to the nature of the theme described and whether it is a marker of quality of life or an attribute that will affect the individual's quality of life. Thus, although one interviewee described how professional work interest was an *indicator* of his personal happiness, it is likely that this is a *causal* variable for his level of quality of life.²⁵⁷ Similarly, many of the themes described above are causal, in that they may change the level of the latent construct of quality of life. This is particularly clear for some of the activities described, for example, exercise and sport for which respondents were often quite clear about their contribution to quality of life. This gives credence to the view that SEIQoL assesses determinants of quality of life rather than being a direct measure of the construct.⁵⁹ This does not detract from its value in describing how individuals evaluate the contribution of various aspects of their life to its overall quality. Furthermore, it makes it ideally suited to exploring how those various components may be re-valued or otherwise altered over time as individuals respond to changing circumstances (such as health or social problems). Whilst an individual's quality of life may be usefully summarised using a single global measure of an existential state, that tells us little of how that state was achieved and maintained.²²⁴ The individualised approach provides that required illumination.

Positive and negative contributors: A second feature is that some cues and themes were expressed due to their potential to either enhance (e.g. friends or family) or detract (e.g. concerns about money) from quality of life. Indeed, the potential to report patient experience that is positively framed (as opposed to, for example, degrees of disability) has been one of the attractions of the quality of life construct for clinicians and researchers.^{195 220} Individualised approaches such as SEIQoL-DW which do not focus upon health (unlike for example, the Patient Generated Index) is a reminder that human existence is not merely to be valued by the absence of infirmity.^{240 258} It is possible that some cues may be more commonly considered because of contemporary problems (e.g. financial difficulties) rather than as a positive contributor to quality of life (e.g. apparent financial prosperity). Similarly, some cues may be more likely provided because of their positive contribution (e.g. social life).

Absent cues: A third and perhaps related consideration is why some potential cues have not been mentioned. Thus, whilst many respondents described family as one of their cues, a few did not. Furthermore, for those who did nominate family, some specifically described children or parents but most did not refer to their partner. Can it be inferred that family or partner were not important to their quality of life? Or is it more likely that such omissions are due to stability or satisfaction with that cue – that because everything is going well it does not warrant explicit consideration? Such an explanation is at least consistent with a 'gap' theory of quality of life – because there is no gap (at least on that domain) it does not present itself as an issue. Therefore, when a discrepancy exists its presence is noted.²¹⁹ Another explanation could be a form of response bias whereby patients are unwilling to discuss some attributes of their life, for example, due to embarrassment.²¹⁵

The latter two explanations may cast some doubt on the general validity of responses although would have little impact upon conclusions from the current analysis. It is also clear though that many respondents were prepared to discuss sensitive issues such as mental health and family problems and financial anxieties. Nevertheless, the question remains whether there are additional contributors to individual quality of life not being incorporated within

the SEIQoL assessment – is it more than the sum of the scale parts?^{229 231}

Supplementing voluntarily provided cues with direct probing of certain 'expected' domains, would potentially address this issue.

In planning, the sampling framework for the study was considered sufficient to allow category saturation, although was not a primary aim of the design. Accordingly no further interviews were intended should this not turn out to be the case (as may be the case in a grounded theoretical study). Whilst certain cues that might be considered to be important influences upon quality of life (such as partners) were infrequently mentioned, they were still mentioned by a minority. Even if such life domains were not provided as cues, they were often discussed during the course of the interview. The stratified sampling strategy sought to maximise variation and a diverse array of cues was ultimately elicited. Whilst there may well be other potential life domains not identified within the current sample of patients, the resultant categorisation still exceeds to scope of most quality of life evaluations of patients with knee problems (as noted above). Furthermore, the 'lack' of saturation has no implication for the assessment of response shift that follows this chapter.

The relationship between cues and quality of life

How cues (and themes) influence quality of life is a key question. For some cues, the amount of time engaged in that activity is important. Thus, for some people work is important because it is time-consuming, aside from its other attributes (e.g. remuneration). Superficially, some cues may appear relatively trivial (for example, 'shopping') and cue weighting may have confirmed this. However, exploration within the interview of the function and role served by such activities illustrates the importance of looking beyond simple descriptors. Respondents offered less insight for other cues – including 'family', which was almost taken for granted as important. Perhaps the respondents' immersion within that cue and its very familiarity, coupled with the novelty of being questioned on the topic made it more difficult to answer such a question.

Classifying content

Previously SEIQoL content has been classified in a variety of ways, including into existential and non-existential responses.²⁴³ Whilst the thematic framework is presented without an implied hierarchy, it is possible to conceive of its content in such a way. Thus a broad distinction could be made between content that represented basic 'needs' and those that reflected aspirational 'wants'. Thus, what was referred to above as 'enabling health' may fit within the former category, whilst pursuit of sporting interests may form part of the latter. If basic needs are satisfied (because, for example, an individual has no apparent health problems) such cues may not even be discussed and the focus is directed more towards 'wants'. Health (or other) problems may serve to shift focus upon the more basic needs. Although it has been argued that basic needs provide necessary, but not sufficient, conditions for individual quality of life, this may ignore the capacity of individuals to adapt to otherwise detrimental change.²⁵⁹ These issues are central to response shift and are addressed further in chapter eight.

Inter-relationship between cues

Descriptions of how cues contributed to quality of life varied in their level of abstraction, which may partly reflect the cognitive complexity of the task. Some respondents were able to provide detailed and cohesive accounts of the role that cues played in their life. Some provided cues were entirely abstract, such as 'independence', but such considerations in fact may have been underpinning the expression of other cues such as health. More abstract cues may be partly the result of greater reflection by the respondent of the consequences of concrete problems or concerns. Thus, whilst a respondent may have nominated either physical health or independence, their concern may be essentially the same.

Such complexity in the make-up of themes also reflects apparent interdependence between separately nominated quality of life cues. Some cues have a direct impact upon other cues (e.g. being able to work will affect financial security and in turn will influence social activities). In this example, respondents may have nominated all three cues or chosen to provide just

one, which implicitly acts as proxy for the others. Choosing to focus upon one or other cue is integral to response shift re-conceptualisation and is described in chapter eight.

The thematic framework attempts to describe how elements important to quality of life interrelate. However, how one individual conceptualises the relationship between elements of their own quality of life (e.g. *independence* being a component of *health*) may differ from another's view of similar domains (e.g. independence and health as *separate* domains). That there are different conceptualisations of quality of life is the basis of an individualised approach and is acknowledged in the framework. The intention is not to present a classification of precisely mutually exclusive categories and data can therefore, be represented in multiple or overlapping themes.

The relative importance of cues

Although cues were formally weighted in the SEIQoL procedure, relative importance of quality of life themes is not represented in the framework. However, relative and absolute importance of some areas is clear from some of the descriptions. Furthermore, how some areas have come to assume greater (or less) importance is also evident (for example, patient #24 describing the increased importance of his family following his knee problems or patient #39 discussing how work has assumed significance with changes in family role). Such changes result both from knee problems and also other aspects of life, some of which may simply be regarded as part of normal development and maturation (e.g. a shifting focus upon education). This starts to evidence the dynamism of quality of life which is directly assessed in subsequent chapters. Some themes (cues) may appear trivial, and indeed may have been perceived as such by the individual. However, even a modest cue label may represent significant value for an individual and serve as a proxy for various concerns and interests. The value of interview probing to uncover such concerns was clear.

Methodological comments

Some respondents found consideration of quality of life problematic. Thus, respondent #38 suggested that she had never given it much consideration before and furthermore, appeared to equate it with suffering. Although the administration of SEIQoL-DW provides orientation for respondents, the facility for individuals to interpret and respond according to their own perspective is a strength of the method. Advantages of the individualised approach includes using weights with personally relevant cues to reflect individual concerns, something that is consistent with a good doctor-patient relationship.^{195 215}

In some instances, further exploration of a cue was problematic (for example, the description of 'family' may not have revealed any further insight into what aspects of family life were important to the individual). Usually this was not the case and such labels were revealed as multi-faceted or representing particular attributes of an overarching theme. Using a qualitative interview to explore cues with respondents provided an extended and contextualised understanding of quality of life content. This enhanced the validity of the resulting framework, and moved beyond a reliance on cue labels and simple cue descriptions to interpret the data - in some ways analogous to cognitive interviewing.²⁶⁰ The resultant insights about cue complexity and inter-relationships are a pre-requisite to understanding the changes that may have occurred over time.

The comprehensive thematic analysis incorporated qualitative descriptive data from the interviews. This extends the categorical analysis of SEIQoL cue labels. The semi-structured nature of the full interview allowed a greater exploration of the cue content. The quality of life domain represented by the cue was not only discussed directly at point of elicitation, but also at subsequent points of the interview. In some instances this merely served to confirm the nature of the cue and provided label. On other occasions, a more refined understanding of the provided cue was possible and enabled clearer concept definition. For some initially provided cues, it was apparent that an interviewee's response may have included more than one salient construct. Interviewing often enabled these separate constructs to be distinguished. For

example, whereas 'work' may have been the provided cue label, what was important for the individual may have been both professional satisfaction and work-related social opportunities. Consequently, a greater range of themes were identified in the thematic framework which, therefore, incorporates the definition of sub-themes within major headings.

The unique perspective of each respondent in describing quality of life content and in agreeing with the interviewer cue labels reveals subtle variations in the meaning that may be attached to any single global theme header. This is not a concern in the current analysis where the content of cue descriptions was effectively extracted to explore and describe themes. It does though indicate the challenge of aggregating data on the basis of cue labels, even when accompanied with short descriptions as would be the case in routine SEIQoL assessments. This issue is explored further in chapter eleven.

7.5 Summary

The expanded quality of life framework has contributed to the response shift research theme of the thesis in two ways. Firstly, it has provided a clearer and more detailed picture of what determined and defined quality of life for this sample of respondents. It has thus simply extended the individualised approach of SEIQoL-DW and exploited the extra time in the assessment and the use of recorded audio data. It has provided a baseline understanding of the complex and idiosyncratic nature of quality of life experience. It has provided insights into the concept of quality of life as experienced and reported by individuals which may be explored and assessed more broadly. Secondly, an extended and more finely graded classification has provided a greater opportunity to explore re-prioritisation and re-conceptualisation response shift across assessment time-points. This means that simple comparison of SEIQoL cue labels across time-points can be expanded to a more detailed assessment of change within and between cue headers. As mentioned above, the thematic framework is not expected to be exhaustive in content as that was not the intention of the theoretical sampling. It is not therefore presented as such, but rather as an insight into the construction and

complexity of quality of life for this study sample. This process has been enhanced by explicit consideration of response shift in the design, conduct and analysis of the data.⁹⁵

Preface to chapter 8

The following chapter presents the main qualitative analysis of re-prioritisation and re-conceptualisation response shift of this thesis. It draws upon the insights obtained from the analysis in the current chapter and utilises a case study based approach.

Chapter 8: Stability and change in quality of life: qualitative evidence for re-conceptualisation and re-prioritisation response shift from individualised assessment

Abstract: *The main results of the qualitative analysis from the interview study are presented within the three sections of this chapter. Firstly, evidence of response shift is described through a series of individual case studies centred upon the expanded SEIQoL-DW structure. In particular, two case studies have been chosen to represent each of three levels of stability and change identified from the analysis. The three levels are (i) stability of cue content and weight; (ii) change of cue weights only; and (iii) change of cue content. Briefer case studies are presented for all study patients in appendix IX. Within each case study evidence for re-prioritisation and re-conceptualisation is assessed. In the second section, conceptually distinct forms of change have been identified from the analysis and presented as a typology. How these forms of change relate to current conceptualisations of response shift change is assessed. The third section focuses upon mechanisms that may facilitate response shift. In particular, this addresses evidence of coping; adaptation; loss; and concomitant change. Each section contains a brief discussion, whilst key messages are reviewed within an integrated conclusion section at the close of the chapter.*

8.1 Comparing cue profiles – a case study approach

The primary aims of the interview study were to determine whether response shift occurs and, in doing so, to explore evidence of re-prioritisation and re-conceptualisation. Changes in SEIQoL-DW cue weight across assessments would indicate re-prioritisation whilst change in cue content would indicate re-conceptualisation. Therefore, cue profiles at baseline and follow-up for each participant were compared. For three participants this was not possible due to a failure to follow-up with a second interview (table 8.1.1). Furthermore, for two patients, cue labels / descriptions were not clearly defined at either

interview. Four patients provided less than five cues at one assessment and one patient was only able to provide cue labels (and not weights) for her follow-up assessment. However, for the latter group of five patients, comparison was still possible.

Table 8.1.1 Reasons for incomplete cue profile

Reason	Patients (n)	Patient id	Comparison possible
Patient not followed up	3	26, 33, 37	No
Cues not sufficiently clarified	2	25, 34	No
Less than five cues elicited	4	16, 17, 21, 29	Yes
Cue label only identified	1	28	Yes

In analysis, three theoretical levels of change were distinguished and described – (i) patients who presented a stable profile in terms of cue content and cue weight, (ii) patients with stable cue content but a change in cue weights and (iii) patients with substantive change in cue content. In the latter group there may also have been changes in assigned weights in otherwise stable cues. However, an underlying change in cue content may have affected the validity of comparing weights. Therefore, the focus here is on the cue content.

Presentation and examples of case studies: For each patient, the level and weight of SEIQoL cues nominated by the patient at baseline and follow-up interviews were summarised using histograms and pie charts respectively. A full summary of all intra-patient comparisons is provided in appendix IX. Within this chapter section, six case studies are presented to represent the three levels of change identified above (two cases per level). Example case studies were chosen firstly on the basis of the main stratifying variable (stage of clinical management). Thus, for each level of change, examples of arthroscopy and waiting list patients are presented (table 8.1.2). Secondly, cases were chosen also to reflect diversity by gender and age where possible. Finally, all cases were chosen as typical and illustrative examples of the level of change being described. The summary SEIQoL index score (potential range from 0 to 100) for each patient is reported for both baseline and follow-

up assessment. The impact of changing cue weight upon index score was assessed by calculating the summary score based upon follow-up cue levels and using both follow-up and baseline cue weights (where cue content remained stable). Extracts accompanying each case study below indicate whether baseline or follow-up interview (I or II respectively) is being used. Cue labels are indicated in the text using italics.

Table 8.1.2 Selected case studies: summary of patient characteristics

Level of change	Patient	Stage of clinical management	Gender	Age
(i) Stable	#15	Waiting list	Female	<40
	#12	Arthroscopy	Female	<40
(ii) Changed weighting	#10	Waiting list	Male	40+
	#5	Arthroscopy	Male	40+
(iii) Changed content	#39	Waiting list	Female	40+
	#30	Arthroscopy	Female	40+

(i) Stable cue content and weight

Few patients in the study demonstrated stability in both cue content and weight across assessments. This partly reflects the precision by which the cues were elicited and classified. Verbatim recording and transcription permitted finer classification of cue content as demonstrated in the thematic framework (chapter 7). However, a precise criterion for stability in cue weight does not exist (i.e. how large a change in weight would indicate change). Therefore, it is probably more appropriate to consider a continuum of stability. Nevertheless, five patients were considered under this heading, including three arthroscopy and two waiting list patients.

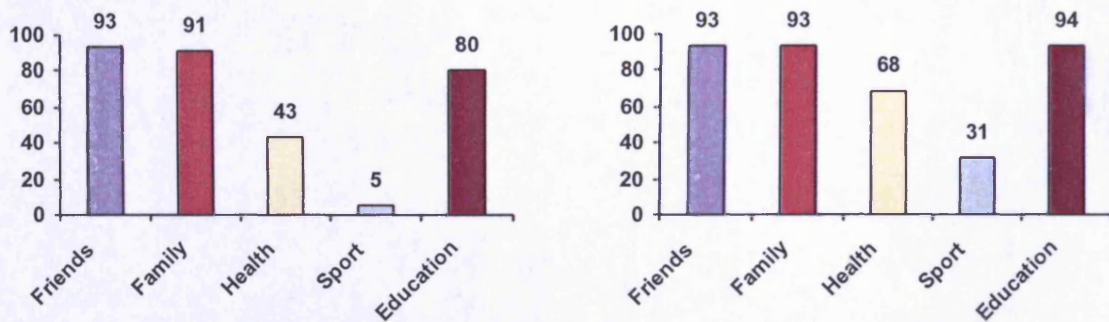
Case study 1: Stable cue content and weight - patient #15

The first example of stable cue content and cue weight was patient #15 (figure 8.1.1). Levels for each cue were also similar across assessments although there is some improvement in *health*, *sport* and *education*. Finally, the weight that she has assigned to each cue has also remained relatively stable. However, there was some decrease in the weight accorded to *health* (10 point difference) and an increase for *friends* (7 points). The SEIQoL index score

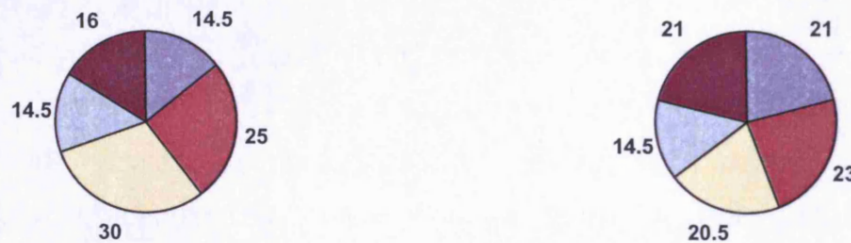
increased from 62.7 to 79.1. When follow-up SEIQoL index score was calculated using the baseline (rather than follow-up) weights, the derived score is very similar (76.7). Hence, any change in weights has had minimal impact upon overall quality of life score.

Figure 8.1.1: Stable cue profile - patient #15 (waiting list / female / <40)
Baseline (index score = 62.66) **Follow-up** (index score = 79.10)

Levels



Weights



In describing her first cue, *friends*, their supportive function was mentioned at baseline and follow-up. Also referred to in the second interview was the entertainment value of friends. Similarly, for *health*, the importance of health as a pre-requisite for the rest of life's activities was consistently described. With regards to *sport*, she described in the baseline interview how this provided a balance to her educational activities. Being unable to participate fully was consequently problematic. *Sport* also encompassed watching sport (pursued as an activity with her father). The team spirit inherit in some participative sports was important for her, as was the competition involved.

She expressed these various sentiments in both interviews. Thus, the patients' accompanying description of cue content supported the stability apparent when simply comparing cue profiles.

The patient perceived the stability in her perspective at the start of the second interview, although this may have reflected a mild reluctance to engage with the interview process:

II#15: Have you got my answers from last time because they're pretty much the same I think?

Her attribution of cue weights was also consistent. She commented about her weighting for *Sport* suggesting that this had already changed some time ago as a reaction to her decreased function:

II#15: ... it's not a big part of my life at the moment because obviously I can't do it. So maybe if it was, then it might go a bit above - I don't know - education or health or something, but family and friends would always come at the top.

Thus, re-prioritisation may have already occurred for this patient prior to baseline assessment, which was some considerable time after her initial injury. That her knee symptoms had not improved during the period between assessments may explain why the weighting for *sports* remained both stable and low. In contrast, the weight accorded to *health* had decreased by 9.5 points from the baseline assessment when it had been the highest ranked cue. When asked about this in the second interview, the patient wasn't sure why she had done this. She suggested that in the absence of good health, *family* and *friends* would (still) play an important supportive function:

MR: Do you think there's any reason why that might be the case, that you're rating health now slightly less than before?

II#15: Um, I don't know I'm thinking that, yeah, 'cos I mean if you haven't got good health then your family and friends are just there to support you so [unclear] even if your health was bad

MR: Mm

II#15: So if, yeah... I don't know why I put health top last time.

MR: Right.

II#15: That's why I think that health is third now anyway.

This is a move on from her expressed view at baseline assessment that good health was pre-requisite for other life activities:

MR: ...why would you say you put health as the most important thing out of these five that you've mentioned?

I#15: Um, cos you've got to be healthy haven't you, cos otherwise you don't have any quality of life at all.

Therefore, even though the cue content remained stable, even for this patient there is an indication that her priorities have started to alter as she re-appraises her life. Her weighting of *friends* increased by 6.5 points from baseline when it had been the lowest ranked cue. At follow-up it was ranked equal second. She felt that this was due to simply knowing her current friends for longer and better, and perhaps a shift in her view of their relative importance to her:

II#15: Um, well I'm closer to my friends now because I've been living with them for three years now.

MR: Right.

I#15: So we have got a closer relationship now. And I suppose um, I, yeah, education is important but it's not the be-all-and-end-all is it to be honest? And same with sport. Yeah, I mean if you haven't got friends you're pretty lonely aren't you?

MR: But you think that possibly you're just a bit closer to the friends you've got?

II#15: That's probably why I put, yeah.

Case study 2: Stable cue content and weight - Patient #12

A second example of stable cue content and cue weight was patient #12 (figure 8.1.2). The patient used alternative descriptive labels at the two assessments to refer to her employment (*work* and *job*) although described the same content within both. Similarly, when she referred to *family* at follow-up (which she had not provided as a label as baseline) she included within this definition her partner (whom she had mentioned previously). She may therefore be employing a somewhat expanded definition of this cue at follow-up. When asked specifically about why she had changed from *partner* to

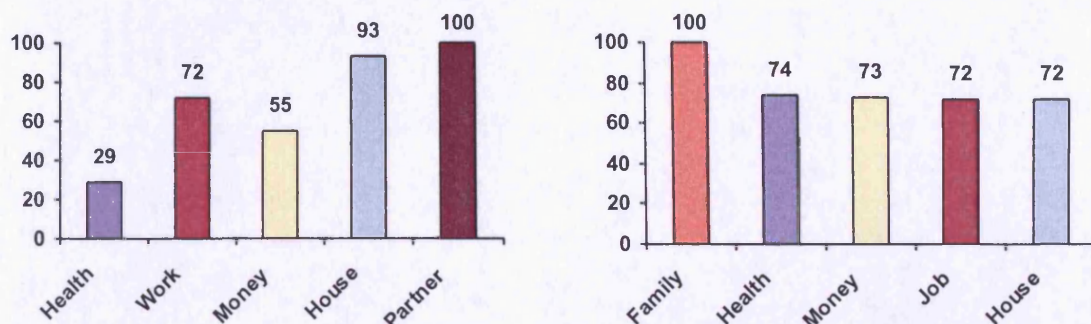
family, she was unable to explain, other than the possibility that he may have been away at baseline.

Figure 8.1.2: Stable cue profile - patient #12 (arthroscopy / female / <40)

Baseline (index score = 60.08)

Follow-up (index score = 81.36)

Levels



Weights



Some cue levels had changed substantially between assessments with *health* improving by 45 points, *money* by 18 and *house* decreasing by 21. Cue weights varied very little over time with the biggest difference being for *money* which decreased by 5 points. SEIQoL index scores were 68.3 at baseline improving to 81.4 at follow-up. Again, when follow-up SEIQoL index score was calculated using baseline weights, the derived score was very similar (80.4).

There was no evidence that there had been a change in her conception of quality of life construction; cue content and weighting appeared stable. Her description of cues appeared consistent. For example, at baseline she described the importance of health for enabling life's activities. At follow-up she provides a similar description:

- MR: *... what is it that you're thinking of when you talk about health and ...?*
- II#12: *Being able to do things without having to worry.*
- MR: *Right.*
- II#12: *No restriction in what I do.*
- MR: *And is that the case at the moment then - that there are restrictions in what you're able to do?*
- II#12: *Um.*
- MR: *Or is it just something that you're....*
- II#12: *There are some, but it's been a lot better since the last time we spoke.*
- MR: *Right. So you're talking about your knee specifically there?*
- II#12: *Yeah the one knee that I've had operated on has been brilliant.*

Despite reported improvement in her knee following re-constructive surgery, there was no apparent response shift. This may have been partly due to the arrival of additional symptoms in her other knee, about which she was expressing caution. Furthermore, her apparent stability may be explained by the excessively long history of knee problems (fourteen years including all her adult life). Accommodation may well have occurred long before the baseline interview:

- II#12: *I'd just forget about it at the time like. I think I got used to it, that's the problem, I'm getting used to this one now. The pain that comes you get used to it and just carry on with it, I don't want to leave it too late otherwise I'm going to end up stop doing things again.*

(ii) Changed cue weights

Case study 3: Changed cue weights - Patient #10

The first example of stable cue content, but changed cue weight was provided by patient #10 (figure 8.1.3). Indeed, at follow-up the patient spontaneously remarked that his priorities had changed since baseline assessment, and that

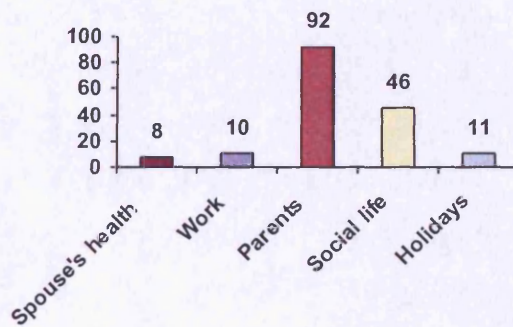
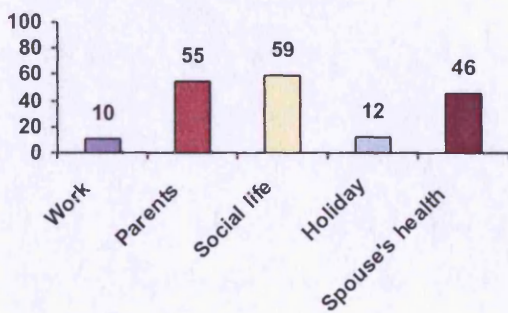
the *health of his wife* was now his primary concern. This assertion was consistent with cue weights at both assessments. The weight assigned to his wife's health had increased by 21 points, whilst that attached to work had decreased by 17.5 points. SEIQoL index scores at baseline and follow-up were low and relatively stable (28.93 and 26.31 respectively). When follow-up SEIQoL index score was calculated using baseline weights, the derived score was 32.03, a difference of 5.72.

Figure 8.1.3 Changed cue profile (weight) – patient #10 (waiting list / male / 40+)

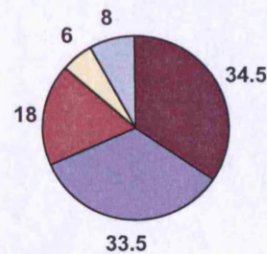
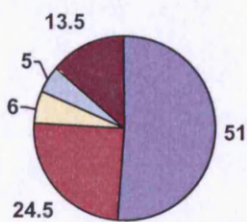
Baseline (index score = 28.83)

Follow-up (index score = 26.31)

Levels



Weights



The patient attributed his changed priority to a significant deterioration in his wife's health. In the first interview he reported his wife was unwell with a chest infection, and mentioned that she suffered from arthritis and depression.

However, at the follow-up interview she had been in hospital for over four weeks suffering from alcohol-related liver damage:

II#10: She didn't drink spirits or wine, it was just a case of drinking all day, lager, beer.

MR: Mm.

II#10: Consequences - she's lost her balance. Short-term memory is getting mixed up with her long term memory.

MR: Oh right.

II#10: She's still saying strange things to me that she thinks has happened. She thinks she's gone shopping and she's not, and she only thinks that she's been in hospital a few days. She doesn't realise she's actually been there for a month like, you know. Of course, she became incontinent solely because she had an infection in the water and over the past few days since they found that out, that's got a little bit better so it was, I was forever bringing home three or four changes of bed wear for her every day.

Whilst the weighting of work decreased from the baseline interview, at follow-up the patient remarked how both his work situation and his wife's health were his main priority. He was still off work despite having attempted to return. His expectations regarding his treatment and the course of his condition were interesting in this regard. At baseline he assumed that he would not be waiting long (e.g. several weeks) to be seen in secondary care. At follow-up, his only formal contact with secondary care had been a letter enquiring whether he wished to continue to pursue his referral. Although he had revised his expectation for waiting time, he remained unaware of the likely delay (potentially two to three years). If this had been his expectation the weight attached to work may well have increased. Nevertheless, he remained frustrated by being continuously signed off work. His inability to work was causing him financial difficulties and he had concerns about his long-term security:

II#10: I think I went into work Monday, Tuesday, Wednesday. Thursday it was bad, I went back into work the following week and the same thing happened again. I done another couple of days and my leg started to give way and I said "that's it, I can't".

MR: Right, so you were indicating to your GP that you wanted to go back?

II#10: Well I wanted to go back solely because I'm running out of money.

MR: Yeah.

II#10: Apart from that, I'm worried about my job.
MR: Yeah.
II#10: You know, they say that they can't get rid of you if you're on the sick, but if it's going to be a long term thing I think they can.
MR: Mm.
II#10: I don't want that to happen, I've been there 13 years.
MR: Mm.
II#10: There's no sentiment in business. They can always replace me with somebody else. Probably somebody cheaper as well ...

At follow-up when asked to comment upon whether he had changed his priorities, he considered that he would have rated his parents higher previously compared to his wife. This was consistent with his actual SEIQoL weightings. Whilst his wife's deteriorated condition was a factor in this, he also indicated that as far as his parents were concerned they were content and that there was nothing further required:

II#10: But they're quite happy with how things are at the moment. So am I at the moment. I mean it can't get any better.
MR: Oh right.
II#10: I know that because of their age.
MR: But you're now living with them whereas before I came to see you were ...
II#10: Yeah, I always wanted to be close to them you see in case things do go wrong.

Since the baseline interview he and his wife had moved into his parents' house so he was now better able to care for them. Whilst this was one motivation for moving, he also admitted that he had also been concerned about being able to pay the rental on his former flat.

Case study 4: Changed cue weights - Patient #5

A second example of stable cue content, but changed cue weight was provided by patient #5 (figure 8.1.4). Change was most apparent for the cues *health* and *marriage*. At baseline, *health* was the most highly ranked cue with half of all the available weight. This had reduced at follow-up and its place had effectively been taken by *marriage*. When asked to comment upon this change, the patient described how his health status was much improved,

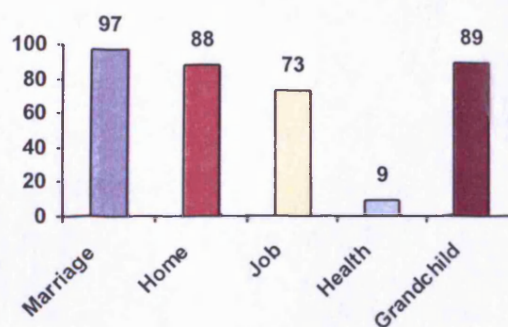
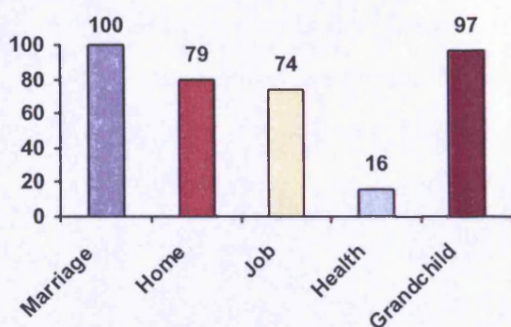
including his mobility, which had been a matter of great concern to him at baseline.

Table 8.1.4 Changed cue profile (weight) – patient #5 (arthroscopy / male / 40+)

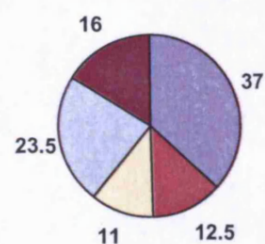
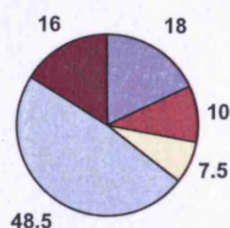
Baseline (index score = 54.73)

Follow-up (index score = 71.28)

Levels



Weights



His reported health improvement was reflected by an increased SEIQoL index score (54.73 at baseline, 71.28 at follow-up) and EQ-5D scores (initially, 0.02 and 0.52 subsequently). Calculating follow-up SEIQoL index score using baseline weights resulted in a score of 50.34 - a difference of 20.95. The patient commented upon the changes that had occurred:

II#5: I'm not such a miserable so-and-so if you like. I mean the marriage I've got is brilliant in my eyes - just the best. And because I can move around, it's improved that as well. I would say that. Health wise you

know, I could always put up with pain. I couldn't put up with losing [name of wife] if you see what I mean. I would hate for my marriage to go wrong.

MR: *Mm.*

II#5: *And I know that she feels the same so that's why I swapped the two over - purely because I'm happier because of my knees.*

MR: *So when I say that the last time I saw you, you said that health was the most important thing at the time, but it doesn't surprise you that looking back at that why you said it?*

II#5: *I am slightly surprised that I said that but I can remove the reasoning behind it, my health at the time wasn't that good and it was affecting everything. It was affecting everything across the board.*

At follow-up he placed greater priority on his wife, although it is not clear whether he had taken a positive step towards this viewpoint or whether it was secondary to the improvement in health. Whether he would have reported a shift in his values in the presence of continuing health problems is unclear. This uncertainty is perhaps reflected in a subsequent comment:

II#5: *Now I've got a better general moods and being happier I've considered there's more to life than health but saying that I'd hate to be unwell. Saying that, I'd hate to have any serious illness.*

It may also be the case that his reported conceptualisation of health may have changed over the previous six months. At baseline he described the significant impact of his knee and (to a lesser extent) his other health problems upon his physical mobility and his mental health. He discussed how he felt old compared to his peers due to their continued enjoyment of sporting and other physical activities, whilst he endured restrictions. He was currently on anti-depressant medication due to the effect that this was having upon him. However, at follow-up he was more focused upon physical fitness and a positive approach to 'getting fit' and less upon the disabling effects of ill-health:

II#5: *For a while afterwards it was brilliant. It didn't tend to swell but I think maybe it's a combination of I'm doing it, so ...*

MR: *So you pushed yourself a bit more then.*

II#5: *Yeah. I'm definitely pushing the amount I walk, the distance I walk and how often I walk as a way of getting fit.*

It maybe that he had been buoyed by feedback from his surgeon who felt his knee was, upon examination, not as bad as may have been expected given his sporting history. Along with the actual improvement in mobility and reduction in pain, he may have felt that things were not as bad as they could have been. In summary, whilst there may be some re-conceptualisation of the health cue itself, the most apparent shifts are in weight accorded to health and marriage. These two cues have remained relatively stable in level despite an improvement in knee function and pain following bilateral arthroscopies.

(iii) Changed cue content

Case study 5: Changed cue content - patient #39

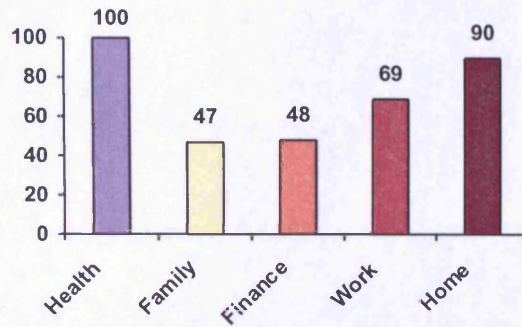
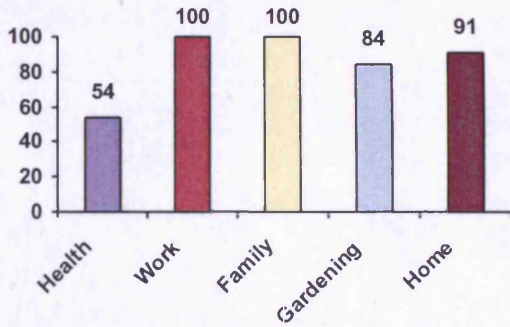
The first example of changed cue content was provided by patient #39 (figure 8.1.5). Her overall SEIQoL index score was 77.5 at baseline but had decreased to 53.56 at follow-up. Four cue labels were consistently nominated at both assessments: *health*, *work*, *family* and *home*. The fifth baseline cue was *gardening* and at follow-up was *finance*.

Figure 8.1.5 Changed cue profile (content) – patient #39 (waiting list / female / 40+)

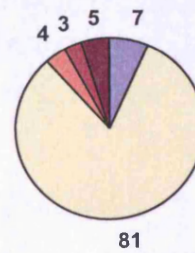
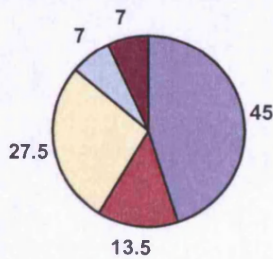
Baseline (index score = 77.5)

Follow-up (index score = 53.56)

Levels



Weights



The first baseline cue provided was *health* which largely reflected concomitant physical health problems (i.e. fibromyalgia). Subsequently, *health* was still a cue label, although now she emphasised both mental and physical health:

II#39: *My health, mental and physical.*

The reasons for this expansion in scope were the considerable problems that she had suffered in the interval between interviews:

II#39: *I've just been off work for nearly four months with stress.*

MR: *Right.*

- II#39: So until it really hits you and you come apart you don't really know. Do you know what I mean - you know that maybe some people are worriers and others aren't ...*
- MR: Would you say - ?*
- II#39: A bit of a wreck really.*
- MR: And you've only just gone back as such?*
- II#39: I've been back now about five weeks I think – four weeks, five weeks.*
- MR: Right. So the mental side of your health has been to the forefront most recently then?*
- II#39: Yeah and coping with its - my knee is really - is not the biggest issue in my health, do you know what I mean? So it's really being able to cope with the other thing more. So you know - and when you're having a good day you just want to get everything done on that day because the next day you might not -*

The impact of her fibromyalgia had been intermittent but the stress had been constant. She suggested, therefore, that she now had an expanded appreciation of the potential impact and importance of mental health problems. Whilst *health* remained constant as an overarching cue, it appears that what she included under this heading had expanded due to her experience – an example of re-conceptualisation.

The relative reduction in weighting for health appears surprising given her significant health problems. However, she felt that her knee had been worse at baseline and that this had been influencing her weighting then. Furthermore, and maybe more significantly, there had been serious family issues apparent at follow-up affecting her cue weighting. Both her adult daughters had ongoing financial problems, and one also had serious mental health problems. The latter daughter wanted to move back in with her parents but the patient was worried that this could seriously damage her own marriage. This precarious balance between partner and daughter explains the low rating for the *family* and why its weight has increased. Her family was also something that she felt she should be able to influence, not necessarily the case for other areas of her life:

- II#39: ... So I'm waiting for a time to come when - I'm hoping I'll never have to make a decision but then it's always hanging there, do you know what I mean? So - and once you make a wrong decision like that I think you can't turn it back. You can't make everything okay*

afterwards and she's sort of tried to commit suicide three times. So that's in your head and but then I think she's very spoilt and very - has to have her own way. So I don't know they just really are important to me. Very - a house is a house and it's made of bricks. Your work - you just get another job. Money comes and goes anyway unless you're rich or very poor, you just plod on with that. Health, if you haven't got it as you would like it, you've just got to get on with it as best you can. But maybe family is, you know, is something you can maybe do something about. So, maybe its because I'm a woman as well and maybe women are different to men and maybe I'm soft [laughs].

Whilst work was elicited at both interviews she described how going to work was important to her, rather than the actual job, which she didn't particularly enjoy. She described the importance of the work role. Once her daughters had left home, she had tried to re-establish a meaningful role for herself in life. Work was one way of achieving this and again represented a change in her quality of life domains, albeit of slightly longer standing than the study period. She returned to this subject in the follow-up interview:

II#39.pc: If you're a certain person, your role is to bring your children up and bring them up as best you can and then when they grow up and leave home you've lost that kind of role. So you replace it with another role and for me it was to going to work full-time. You know, a tidy job as opposed to part-time jobs, so you give it 120% or 130%.

MR: Right.

II#39.pc: So, work is important because you are there more than you are at home but it's only a job at the end of the day and if that one ends then you find something else.

MR: And what's – and what's made you come to that conclusion for yourself?

II#39.pc: Because I was ill.

MR: Yeah. With the stress do you mean?

II#39.pc: Yeah.

Her illness has served to emphasise this concept of work and that any individual job was less important than simply being employed. She reported being less anxious about work than she used to be, and commented that being ill with stress was due to her worrying about work.

The cue mentioned only at baseline was *gardening*. At follow-up, she concluded that she had not mentioned this cue as it was currently the wrong time of the year (November). The baseline interview was in April and thus cue provision may have been seasonally determined. In contrast, the cue mentioned only at follow-up was *finance*. The patient put this down to the financial difficulties experienced by both daughters. This had been a chronic problem with one daughter. However, she had only recently discovered that her other daughter also had financial problems. This had come as a 'complete shock' to her. Hence this apparent change in circumstances had effected a change in cue content.

For this patient, subtle changes within the content of single cues (an expanded scope for health), changes in cue weight (an important increase for *family*, a reduction in all others including *work*) were accompanied by the introduction of a new cue, *finance*. A baseline cue was possibly only seasonally important and most changes were recognisable and explained by the patient themselves. In her broader life context, the influence of the knee problem was apparent but relatively small.

Case study 6: Changed cue content - patient #30

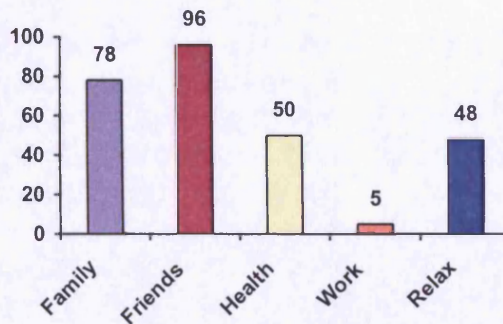
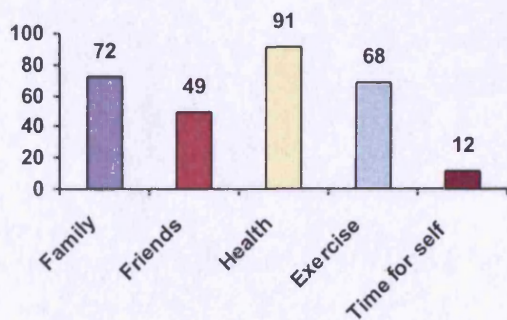
A second example of changed cue content was provided by patient #30 (figure 8.1.6). Her SEIQoL index score remained relatively stable across assessments – 63.09 at baseline and 65 at follow-up. Three cues were consistently labelled across assessments: *family*, *friends* and *health*. A further baseline cue was *time for self* and described similar concepts to that included within the follow-up cue 'relaxation'. The baseline cue, 'exercise', was not mentioned specifically at follow-up, whilst one cue was only mentioned at follow-up - *work*.

Figure 8.1.6 Changed cue profile (content) – patient #30 (arthroscopy / female / 40+)

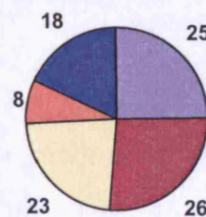
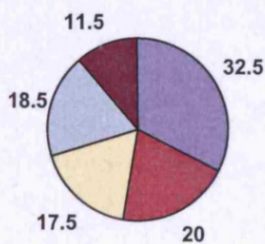
Baseline (index score = 63.09)

Follow-up (index score = 65.00)

Levels



Weights



Whilst her description of the 'family' cue was consistent across interviews, it had already increased in importance following recent family bereavements:

Il#30: A close family, and I mean sort of losing my parents and my husband's parents in the last 10 years makes you realise that brothers and sisters and that extended family just that little bit more important to you, you know.

At baseline she discussed the nature and role of a small group of close friends. Subsequently, she described how she had come to place special

emphasis upon her friends, primarily following recent problems in work (she was suspended pending a disciplinary inquiry and was ultimately vindicated). Throughout this time she was disappointed by the support she received from certain friends. In contrast, she also received sometimes unexpected support from others (including work colleagues and acquaintances). She described her feelings:

II#30: Friends - I've realised this year how important. I always thought friends were really, really important but I've got three real close friends that have sort of stuck with you through thick and thin. Um, and lots of other friends you say, "Oh yes, it's a friend", you realise become more like colleagues or acquaintances.

MR: Mm.

II#30: And some really good colleagues who have turned out to be good friends, you know. Stuck by me this year.

Thus, the basic role and function of friends remained stable but the importance attached had increased, as evidenced by a higher SEIQoL weight at follow-up (now the highest ranked cue). Who she counted amongst her friends had also changed during the intervening period.

The third stable cue label was *health*, which at baseline was mainly focused upon her knee problem and initial concerns prior to achieving a diagnosis. At follow-up, her deteriorated health had consequently taken on greater significance:

II#30: ... Um, health probably comes now further up the line.

MR: Right.

...

II#30: Yeah. I mean then at the time I thought I seemed to be doing pretty well, health-wise.

MR: Right.

II#30: But since then I've sort of gone downhill in the way that I've put on a couple of stone and I'm not quite as fit as I was.

MR: Right.

II#30: That needs to come at the top of the line for me now to give me better quality of life ...

Thus, her increased weight was a problem and, at follow-up, she also disclosed work-related problems with stress and consequent use of anti-

depressant medication. Therefore, the level for health was lower at follow-up and her description of the increased importance attached to health was reflected by increased SEIQoL weighting at follow-up.

'Exercise' was only elicited at baseline and she described how she was practically addicted to it. She had been heavily motivated by medical advice regarding possible long-term consequences of her knee problem. Exercising had reduced both her weight and the impact of her knee problem to the extent that she opted not to pursue arthroscopy. Subsequently though she had stopped exercising and increased her consumption, due to her work problems which had left her de-motivated. At the time of the second interview, she had just returned to exercising at the gym:

I#30: Well I mean I got to the point a couple of years ago when I was still heavy and I was sort of relying on a walking stick virtually all the time, that it couldn't go on like this. I hated the thought that it was going to make me... as the doctor said if you don't lose weight you'll end up in a wheelchair.

...

I#30: And um, I think it's probably made me realise that problems can take me to food. I never thought before I had a problem with the fact that - I've seen in the years my mum being a great worrier and whenever she was worried about anything or concerned she'd starve herself and not want to eat you know, I got the opposite and ate everything in sight.

At baseline she described the importance to her of having time to herself (a response to her perception that she habitually prioritised her family and work). She achieved this by going to the gym - an example of the close interrelationship between elicited cues. In this instance, one cue (*exercise*) served multiple roles including a means of achieving another cue (*time for self*):

I#30: Yes, time to myself with the exercise. People might say that if you can get to the gym 3 times a week you're lucky you get time for yourself, but to me that's a necessity.

MR: And that's one of the few times you feel you can have time to yourself.

I#30: Yes, I like the actual total switching off and nobody can contact me. I've got my mobile switched off locked in my locker. I usually try to

time it so that if there's something on telly that I can actually find ok and I'll watch it and if, like this morning, I would have usually gone to the gym and watched Lorraine Kelly first thing this morning but if I was home I wouldn't be, I'd be hoovering and so I pin-point times like that and just switch off.

One of the consequences of exercising at the gym and having this time for herself was relaxation. Thus the themes of relaxation she describes as a cue at follow-up is integral to the 'time to self' she describes at baseline:

l#30.ph: I just go on the cross trainer and the treadmill. I can't run I still can't run even though I've lost 6 stone I still can't run, but I just spend time watching the screens up there and treading away for a few miles whereas I wouldn't think of walking round the streets. Car driving in the car, and it's sort of my time you know. With a 13 year old I don't seem to get much time that's mine so that's my relaxation.

MR: So you get a number of things out of the exercise really? ...

l#30.ph: I like to do it on my own. I don't like to go with friends. That's the sort of thing that I don't you know. It takes it away from what you're doing. I like to go there stick on the headphones and that's my 2 hours. A shower at the end of it and then when I come home I just feel so much better for it.

The cue only provided at follow-up was *work*, although she assumed that she had mentioned this at baseline. In fact she had discussed her work at baseline, but not nominated as a cue. She described how she wanted to reduce work commitments and work part-time. At baseline she commented that work would previously have been a quality of life cue but that it had become more as a means to an end. At the second interview she thought she had now included *work* because of her recent work difficulties. She considered that it wasn't a cue previously because she simply accepted work as a necessary and commonplace activity.

In summary for this patient, certain cues had assumed greater importance over time, including *friends* and *health*. Another cue mentioned at baseline, *exercise* had dropped from her cue profile, whilst another had been introduced (*work*). She was able to account for these changes which were in large part a response to her negative experiences within her work. Her health

concerns included her knee but also encompassed other factors such as stress and weight.

Comparing cue profiles – a discussion

The case studies provide evidence that both re-prioritisation and re-conceptualisation response shift have occurred in this sample of patients. The modified SEIQoL-DW schedule with greater probing of respondents, and a qualitative analytic approach provided greater insight into the nature and mechanisms of change. A case study approach which placed cues in a broader life context and sought descriptive evidence of apparent change served to support the credibility of conclusions about change.

Apparent and real change: Simple comparison of cue labels at different assessment time-points to determine stability of change may be misleading, especially if the accompanying cue description is brief. The analysis of cue content in chapter 7 and the case study approach has helped to clarify where change is likely to be important and meaningful for the respondent and where it merely reflects a difference in how a respondent expresses themselves. Direct probing of cue content and weight changes, a form of respondent validation, was particularly helpful in this. A consideration of different forms of change apparent in this study is presented in the following section.

Timing of response shift: In some case studies there was a suggestion of response shift occurring some time before the study started (e.g. patients #15, #12, #30, #39). This may have been related to the knee or other factors in a patient's life. This is not surprising since many patients, including those recently referred, actually may have been experiencing knee problems for a considerable time.

Re-prioritisation: Within the case studies presented there was a suggestion that the weighting attached to a cue may be influenced either by the level of a cue or, possibly a change in cue level (e.g. patient #5). Thus as a problem in

one life area resolves, it may lead the weight attached to it to reducing (or vice versa). This subject is addressed further in chapter ten.

Wagner and colleagues observed that for identical life domains different rationales for importance were provided.²⁶¹ This is clear from the case studies where overarching cue labels represent a variety of content, which itself will vary in value to different individuals. For example, different aspects of the health domain may be emphasised by different patients. The case studies also show that the emphasis (and therefore weighting) may change over time for any one individual.

Impact of re-prioritisation upon measured quality of life: Although cue weights may change between assessments, this may have little consequence for SEIQoL-DW index score. This is likely if levels for each cue are relatively similar (for example, all high or all low). However, if there are large differences in levels between cues a change in weight may substantially alter the index score. This is demonstrated in a theoretical example (table 8.1.7). Using the follow-up weights would produce an index score of 48.5, using the baseline weights it would be 67.5 - a difference of nearly twenty points. Thus re-prioritisation may substantially affect outcome measure score, whilst domain content remains stable.

Table 8.1.7 Impact of change in cue weight upon SEIQoL index score

Cue	Cue level	Baseline weight	Follow-up weight
A	90	30	10
B	80	25	15
C	70	20	20
D	30	15	25
E	20	10	30
Index score:		67.5	48.5

Re-prioritisation or re-conceptualisation: Although re-prioritisation and re-conceptualisation have been identified in the case studies separately, they are likely to be inter-related. Lowy and Bernhard considered re-conceptualisation a particular case of change in values, specifically where the weight on one domain reduces to near zero.¹⁵⁷ They refer to both as re-conceptualisation.

Although respondents may be offered up to five cues to nominate, theoretically additional life areas may be identified by individuals as relevant to their life quality. Thus, a highly ranked cue may reduce in importance over time to be left as the fifth ranked SEIQoL cue (re-prioritisation). If its relative importance reduces further – for example, to sixth ranked position – it may still play some role in life quality (albeit less influential). Although, within a SEIQoL-based evaluation framework this would appear as re-conceptualisation, it reflects a re-prioritisation of cue. Whilst a cue may become of no importance to an individual, it is perhaps more likely that it simply just becomes less important.

Rapkin interviewed patients twice over six months to identify changes in personal goals without limiting the number of goals they could provide.¹³⁸ Allowing unlimited cues or using an additional category for 'other' aspects of life (as in the Patient Generated Index) may clarify the changing status of such cues.¹⁸⁶ However, such an approach may present difficulties for respondents, for example, in how to aggregate such data or separate out different areas of life.¹⁹⁸

Permanence of change: Two emergent issues identified by Schwartz and colleagues also arise in the current study, firstly whether response shift represents permanent or temporary change.⁹³ The adaptation process may involve fluctuation rather than simple linear change as patients search for a balance.⁹⁴ Whilst Sprangers and Schwartz viewed response shift as flexible and dynamic, others have assumed response shift to represent permanent change.^{10 110} The case studies show evidence of both transient and more permanent change, both of which may vary in personal importance. Temporary changes, for example, sport or gardening which may be seasonally determined (and not simply random in their expression) will nevertheless affect formal assessment.

Potential catalysts: The second emergent issue was the role of other (non-health) factors as catalysts for response shift. It is clear that other life events influenced change, which makes sense if quality of life is viewed as more than

simply health-related. Even only health related quality of life would be influenced by non-health factors.⁹³ Although patients at different stages of clinical management were included in the study, response shift changes may have occurred in both groups and for reasons other than the knee problem. This emphasises the difficulty of identifying a suitable control group in response shift studies.²⁶²

8.2 A typology of change

Perhaps the most apparent form of change when using SEIQoL to assess response shift is cue content (cues being newly included within, or being omitted from, an individual profile). Thereafter, cue weight changes are apparent when the same cues are nominated at both assessments. However, there were other, more subtle forms of change observed across assessments made possible by the detailed qualitative nature of the interview and analysis. Conceptually distinct changes may occur at both the level of the individual cue and also at the level of the overall cue profile (table 8.2.1). This section introduces the typology of change derived from the case study analysis and reviews the implications for detecting and understanding response shift changes.

Cue-level change

Changes at cue level can occur as a consequence of firstly, cue re-labelling and secondly, cue re-formulation.

Cue re-labelling: As its name implies, cue re-labelling involves the descriptor used by the respondent to name the content of the cue. However, the content being described may not have changed at all, only the label. An apparent response shift change may simply be superficial. An example is provided by patient #30 who used the descriptors 'time for self' and 'relaxing' to refer to similar content at baseline and follow-up interviews respectively. It is possible that the choice of label may represent a shift in focus for the respondent

regardless of content, more akin to a change in weight. In itself though, it may not represent strong evidence of response shift.

Table 8.2.1 A typology of potential response shift change: summary of real and apparent changes in cue profile, cue content and weight

Level of change	Change category	Sub-category	Description and variations
Cue	<i>Re-labelling</i>		i) Stable content but different label used
			ii) Changed content but same label used
	<i>Re-formulation</i>	<i>Expansion</i>	Scope of cue extended (increased definition or specification of cue components)
		<i>Contraction</i>	Scope of cue narrowed
<i>Merger</i>		Originally distinct cues merged and presented as single cue	
	<i>Separation</i>	Original cue separated into two or more components and provided as distinct cue	
Profile	<i>Re-placement</i>		i) New cue added to re-formulated cue profile
			ii) Original cue replaced by new cue

However, cue re-labelling may also mask change if the same descriptive label has been used to describe distinct quality of life content. Such change is only fully revealed by more depth analysis of the elicited content. Analysis of group level aggregate data may not be sufficient to distinguish between such actual and merely apparent changes.

Cue re-formulation: More substantial changes occur at the individual cue level and have been grouped under the heading 'cue re-formulation' which refers to a modification in the constitution of the cue. It thus implies that a cue may be multi-faceted. Four types of cue re-formulation were distinguished *expansion*, *contraction*, *merger* and *separation* although clearly there is some overlap and interdependency between these processes. Cue *expansion* involved the scope of the original cue being extended beyond its original limits. Also included under this header are situations where there is an

increased specification or clarification of the constitution of the cue. An example of expansion is where the baseline cue 'health' was focused primarily upon the knee problem, whereas at follow-up other factors were being explicitly considered, such as hypertension (patient #31) or mental health (patient #39).

Cue *contraction* involves a narrowing of the respondent's focus within their definition of the originally elicited cue. Thus, patient #1 describes her grandson who at baseline is an important contributor to her life quality. At follow-up, the grandson has largely been discounted following accommodation to internal family problems which have minimised their contact. She has narrowed her family focus upon other family members.

Cue *merger* and cue *separation* are more extensive forms of re-formulation and are more easily observed. Merger involves the aggregation of two (or more) originally elicited cues within a combined cue header at the second assessment. For example, patient #23 describes two cues at baseline - 'family activities' and 'marriage'. At follow-up she simply provides only one cue, 'family' which accommodates the content of both original cues. Cue separation is the same process but reversed, with components of an originally provided cue being subsequently presented as separate cues. The same patient (#23) provides a single baseline cue ('being healthy'), whose content at follow-up is described under two cues - 'health' and 'keeping fit'.

Superficially, both cue merger and cue separation represent stability in quality of life conceptualisation as the same content is being described on both occasions, albeit packaged differently. However, merger and separation change the number of cues represented by the same content. This permits or prevents respectively other cues being nominated during cue elicitation. Using an individualised approach such as SEIQoL which limits the number of quality of life domains that can be provided, may lead to erroneous conclusions about re-conceptualisation if data are only analysed at a group level. A 'new' cue at follow-up may well have been an important consideration for the respondent

at baseline, but only subsequently revealed following the merger of two baseline cues.

However, re-formulation of cues by merger or separation may well represent true changes in quality of life conceptualisation. Re-appraisal may well have resulted in the respondent partitioning one domain into separate parts as their insight into how it affects their life has grown. Perhaps more importantly, such re-formulation may well reflect re-prioritisation with the identification of separate cue components representing the greater value being placed on the original cue overall and specifically on certain aspects of it. However, re-formulation changes confound the detection of re-prioritisation using SEIQoL in a grouped level analysis because which cues should be directly compared is unclear.

Profile-level change

Profile replacement: At the level of the overall cue profile further changes are apparent, as predicted by Schwartz and Sprangers.¹¹ This involves cues provided as baseline being omitted from the cue profile at follow-up (and vice versa) and in this study has been termed replacement. There are two ways by which this may occur. Firstly, a re-formulated profile resulting from cue merger creates a vacancy which is filled with a new cue. Secondly, when a cue has been omitted (due to a decrease in absolute or relative importance) it is replaced by a cue that has become of greater importance to the individual. In both cases, the new cue may represent content that was of peripheral importance at baseline (i.e. considered relevant by the respondent but not nominated as a cue) or largely novel (i.e. only newly arising).

Discussion

Masking and revealing change: The typology of change, and the domain framework presented previously in chapter 7, emphasise the complexity of change that may be revealed by repeated SEIQoL administration. It is further clear that the typology represents a clear version of change that is in practice more opaque. Apparent change may not in fact prove substantial. Thus, in a

sample of young people with diabetes, Wagner and colleagues found that choice of cue label may be influenced by context of elicitation.²⁶¹ For example, the label 'diabetes' or 'health' may be provided depending upon whether elicited in diabetes camp or school. Analysis of response shift using individualised assessments may mask change if cue labels and profiles are analysed at only group level. This is methodological challenge which is explored further in chapter ten. In contrast, the extended qualitative application of SEIQoL-DW helps reveal the variations of change expressed within the typology above.

The typology does not necessarily indicate whether the various changes represent response shift. In practice several changes may occur for one individual. For example, expansion, contraction, merger and separation may occur concurrently but the overall content may remain broadly stable. However, such re-formulation may well indicate shifting emphasis for an individual in relation to their quality of life areas. In addition, it may not only be the relative weightings of cues that are important but how the individual sees the interrelationship between them.

8.3 Mechanisms contributing to response shift

A further aim of the interview study was to explore mechanisms contributing towards response shift. This analysis drew upon data from the entire interview including specific probes used in the latter half of the schedule. There were many examples of people trying to maintain their lifestyle and expected behaviour, as well as descriptions of change as a consequence of the knee problem. Emerging from these descriptions were themes of coping, adaptation and accommodation. Also apparent were concepts of loss and the ability of individuals to accept change and move on psychologically in their lives. Each of these elements is described in further detail below.

Coping response

Several patients described their perception of how they had coped with the knee problem, usually in response to an interview probe. Respondents varied and included those who felt that they had coped well and not so well. Even for those in the former group, negative emotions may still have been expressed – such as being unhappy or frustrated. Leaving the interpretation of coping to respondents meant that both physical (including functional) and emotional factors were being considered. Some noted how physically they remained the same but their attitude had changed (e.g. patient #22). The first two patients below describe good coping but define 'coping' differently - being able to continue working and not moaning to his wife respectively. At follow-up, the knee problem for patient #18 had largely resolved itself and for patient #22 had been made bearable by adaptations in work and other areas of his life. The third patient below (#24) had been profoundly affected by his knee problem and was unable to work at either baseline or follow-up.

II#18: Um well I think I coped quite well because I still went to work.

MR: Right, okay.

II#18: You know, I mean I took pain killers.

(Female / 64 years old / waiting list / retired)

II#22: I think I've coped pretty well. My wife doesn't hear me moan about it all the time anymore. Before it was moaning completely about my knee. Now I just don't do nothing so I think I'm coping quite well.

(Male / 34 years old / waiting list / carpenter)

MR: How do you think you've coped in terms of what's happened to your knee?

II#24: Um, very badly really.

(Male / 56 years old / arthroscopy / glazier)

Whilst coping may imply an internal response, the perspective of others was occasionally raised. Indeed, this is the case for patient #22 above, for whom coping is evidenced by what his wife observes. Similarly, for another patient a criteria for coping is the extent to which other people may have noticed the knee's impact upon her:

MR: *How do you think you've sort of adapted or coped with the problem that you've got with your knee in the last few months?*

II#38: *I don't think most people notice it.*

MR: *Right.*

II#38: *In that nobody has said to me "You look as if you're in pain"*

(Female / 47 years old / waiting list / teacher)

For one patient, whose wife was present in both interviews, there was third party evidence of coping which generally confirmed the patient's emotional state but indicated a greater impact than that admitted by the patient:

II#36: *I've just been grumpy, haven't I? That's the bottom line. I can't say anything different, can I?*

Wife: *I'd say he hasn't.*

II#36: *What, been grumpy?*

Wife: *No, I would say you haven't coped ... I would say he's been bordering on depressed ... The kids' joking, "Oh, here he goes, you know, he's tired. He's going to blow".*

(Male / 40 years old / arthroscopy / builder)

Coping involved *acceptance* of the knee problem – that it was a fait accompli, even if that was associated with continuing reservations. Acceptance was evident for both the physical symptoms of the knee problem (e.g. pain) and also the functional limitations it imposed (e.g. it being 'awkward' – patient I#20). That there was no choice involved was described by some (e.g. patient II#21) and others expressed a sense of resignation and passivity (e.g. patient II#15). Others commented about 'getting on with' life and viewing what had happened as 'part of life'. Furthermore, there were also descriptions of how the knee problem had now become part of life; part of the background rather than the foreground. This is demonstrated by the following patient whose discussion of health in the follow-up interview focuses upon migraines, viral infections and smoking. She only includes her knee when prompted:

MR: *Okay, any other things that come to mind when you think about health in relation to yourself?*

II#4: *My leg, but that's just life.*

(Female / 24 years old / arthroscopy / student)

Some patients describe how such acceptance was shaped by earlier life experiences, for example of a serious illness such as cancer. Other comments appeared less grounded in actual events and more a reflection of a personal philosophy, general personality or approach to life. However, the extent to which the knee problem could be accepted as part of life might also be related to the actual and expected duration (i.e. whether short or long-term). The following extracts illustrate these points:

II#4: But I think I just have to get on with it. I try to be level-headed about it anyway. Yeah

(Female / 24 years old / arthroscopy / student)

I#39: But if it's something you've just got to put up with - well, you've got to put up with it and as much as you might not like it, it's called 'Life'.

(Female / 50 years old / waiting list / administrator)

I#35: That time 20 years ago I thought, "God, I've beaten this cancer, what's this arthritis going to be – nothing, nothing." But then, when it makes its impact on you that's when you think about it. But I think to myself, "Well I'm still here, so what have I got to moan about?" You know that's the way I look at it.

(Female / 63 years old / waiting list / retired social worker)

II#22: The pain is constantly there all the time whereas I've learnt to live with that and I'm not expecting to be seen for another year or two so you've got to get on and do the best you can.

(Male / 34 years old / waiting list / carpenter)

The extracts above reflect personality attributes, such as stoicism, which will vary between individuals. For some, a period of coping led up to a point at which action had to be taken, such as consulting their GP or being unable to work or walk. Until such trigger points, inconveniences and discomfort could be tolerated. One patient describes how she delayed consulting her GP hoping that the problem would spontaneously resolve itself. Whilst for patients facing a long wait for effective intervention stoicism may be necessary and adaptive, in this instance such a perspective has been detrimental by delaying intervention:

II#15: Yeah I'm not one of those people who goes to the GP a lot.

MR: Right

II#15: It's a stupid kind of macho thing.

...

II#15: ...It's kind of, "Oh yeah, it'll get better", but it was never going to get better so ...

(Female / 19 years old / waiting list / medical student)

Some described factors that appeared to facilitate their coping, including the use of overt comparisons which included people with either similar conditions; more serious medical conditions; or with worse but non-specific problems. For example:

I#4: I mean there's people far worse off than me – it's just a case of getting on with it then.

II#4: Very well, yeah. It's not as bad as what could be happening is it so.

(Female / 24 years old / arthroscopy / student)

II#39: I mean it really hurts, it really does hurt but then I just think "Well, it's not cancer, it's not life-threatening", and get on with it.

(Female / 50 years old / waiting list / administrator)

II#17: It certainly doesn't make you happy that you can't do them because you want to do them and you feel "Well, why am I missing out?" But you think, "Well, mine's more minor than somebody else's".

(Female / 51 years old / arthroscopy / school support worker)

However, such comparisons also served to inhibit coping if, for example, the patient was reflecting upon their own previous, more healthy state. This included a conception of personal 'normality' - now at odds with their physical reality. That a patient had not amended this self-conception and that there existed a gap between expected and experienced state appeared to inhibit coping:

I#24: As I say I just want to get my life back and if that means me having - going on tablets to get me back to that stage I would do

(Male / 56 years old / arthroscopy / glazier)

Wife (I#36): He's gone on continuously how he's a cripple, he's a cripple. I think he genuinely probably does feel that way as well because he's been so active as well you know. So he feels immobile, unable to cope ... he needs to be up and doing things and getting out and everything else.

(Male / 40 years old / arthroscopy / builder)

The first patient above reports encouragement from his family that things will get better, although if his condition remains stable this may not be helpful if it reinforces unrealistic expectations. Nevertheless, he does appear to have made some concession in their expectation or hope of a return to normality, by conceding that some functions and activities may not be retrieved:

I#24: Um, as I said really it's to get back to normal and do things that, well not all things that I done before, but the majority of things.

(Male / 56 years old / arthroscopy / glazier)

Other factors apparently inhibiting coping included for one patient not being able to exercise (normally a mechanism for coping with general stress).

Physical coping mechanisms mentioned included prescription drugs (painkillers and anti-depressants), tobacco and alcohol. Coping on a 'day to day' basis was also mentioned by one patient (II#1), whilst in contrast another described how the passage of time helped (II#19). However, the former patient continued to experience significant pain in her knee whilst the latter was pain free. One patient described how the longevity of her problem meant that family and friends had also come to accept their knee problem as a normal state of affairs. Others clearly saw that time had allowed them to come to accept their condition. Another patient (patient II#34) described how over time he had accommodated to ongoing pain so that he know was feeling better:

II#34: It haven't improved, I mean I still got that pain there.

...

II#34: What it is now, 'cos its been that long I'm getting used to – what can I say – I'm getting used to me clicking and whatever.

(Male / 46 years old / waiting list / water mains manual worker)

II#14: I do get a bit of pain and discomfort with it now and again but um, it's been so long now since I've had the problem that I just virtually accept the way it is I suppose.

MR: Right, so –

II#14: It's a bit of being resigned to the knee is like it is and just get on and do things.

(Male / 37 years old / arthroscopy / engineer)

II#15: Um, my dad's pretty laid back about stuff like that anyway. Um, cos it's been going on for so long now it's just a matter of course, like. You know, I don't get asked, "How's your knee today?" because it's just the way it is.

(Female / 19 years old / waiting list / medical student)

A more active psychological strategy described by another involved shifting focus towards current physical abilities and away from functional limitations:

II#35: It's what I can still do, what I still want to do. That's it you adjust.

(Female / 63 years old / waiting list / retired social worker)

The same patient also compared her positive approach to that of an elderly relative who she perceived to have 'talked himself into a wheelchair' in the wake of a combination of physical problems:

II#35: Which I found very sad. Never thinking what he could do, but what he couldn't do.

(Female / 63 years old / waiting list / retired social worker)

Not allowing the knee problem to 'rule' one's life was another positively expressed orientation (patient I#16). Similarly outlooks included, for example, belief in future improvements in medical technology if future intervention was required (patient II#17), hoping that the knee won't cause problems (for example, on holidays) and being optimistic in general. Some patients described simply avoiding thinking about future problems, whilst others anticipated and appeared to make concessions to the possibility of deterioration or progression of their problem:

II#17: No, cos I push it to the back of my mind basically.

...

II#17: So what's the point of me sitting here thinking about it all the time. It's going to make me miserable because I just think, "Well, technology's moving – that it's such a wonderful thing", and by the time I need more treatment I'm hoping ...

(Female / 51 years old / arthroscopy / school support worker)

II#20: Yes, it'll probably affect me in the long run ...

...

II#20: If it happens, it happens.

(Male / 29 years old / arthroscopy / security guard)

Adaptation

Whereas coping focused upon acceptance of altered circumstances, adaptation is here used to represent a specific behavioural adjustment. This includes evidence of modification in the pursuit of specific activities; replacement of activities with a substitute; and also the simple cessation of activities.

Modification: A wide range of behaviours were described as having been *modified* by patients, mostly related to general mobility. Such adaptations involved anticipation and, if possible precautions to avoid negative consequences associated with their knee problem. These included specific physical actions such as walking and driving (e.g. taking extra breaks on long trips and sharing driving with a partner). Less dynamic actions were also reported as modified, for example, altered sleeping position. Broader activities were also described which may encompass a number of various physical actions. Such examples included changes within work to avoid kneeling and climbing, changes in domestic activities (e.g. gardening) and changes in sporting activities (e.g. playing less vigorously). Examples of changed actions and activities are described below:

II#38: And I take shorter steps and you know – well that's what you do to compensate don't you?

(Female / 47 years old / waiting list / teacher)

II#32: I guess probably if I'm driving long distance I would tend to think about where I could possibly stop if I had to and go for a walk around until the numbness had gone away in the leg.

(Male / 46 years old / waiting list / IT project manager)

II#22: I've managed it and changed certain ways I spend my time in work and the jobs I take on. If I do flooring I tend to do big gap of flooring and then I won't do any for a week.

(Male / 34 years old / waiting list / carpenter)

Most of these examples are *primary modifications* in that the amended behaviour was directly affected by the knee problem. Changing the way one walks would be an example of that. In addition, *secondary modifications*, involved behaviours being amended which were only indirectly affected by the knee problem. For example, the following patient describes how he is more careful about his diet now that he can no longer exercise to control his weight:

II#5: So that's the only way I can do it really, by cutting down on food and cutting down on beer and trying to get healthier that way cos I know I'm not going to be able to go running to lose the weight.

(Male / 52 years old / arthroscopy / electrician)

Cessation: Some behaviours were not modified, rather they simply ceased - for example, playing a sport. One woman described how she was now unable to kick a ball around on the beach with her young daughter and that she was no longer prepared to engage in certain school-based play activities:

I#17: Um, well as I said, you know, we take the dog down to the beach. We walk, my daughter likes to kick the ball around and I find I can't do those sort of things because my knee would sort of give out on me.

(Female / 51 years old / arthroscopy / school support worker)

Substitution: Occasionally, as above, such behaviours were not replaced. More often though, activities were introduced in their place. This *substitution* was apparent in various areas such as social life, work, sport and also exercise, where for example, a problematic form of exercise could be replaced by something more feasible. Other areas of functioning were also described, such as holidaying (i.e. focusing upon holiday activities which were

less physically demanding). Some examples are provided below, including the first in which the patient describes how his former social life has ended but has been replaced by increased interaction with his extended family:

II#24: You know but – they say “Times a healer”, and that side of it. I think mainly because my children are a lot closer to me now whereas, um, I have four or five phone calls every night now, like “Are you alright dad? Be careful, look after yourself”, and whatever and they call down. But sort of going out skittling and things like that, it’s a thing of the past now.

MR: So you see more of your family and they’re more important in a sense?

II#24: Yes

(Male / 56 years old / arthroscopy / glazier)

II#22: I like, even though I used to do a hell of a lot of running, which I can’t do anymore, I’ve changed now. I’ve started doing dragon boat racing.

...

II#22: So that’s something I can do with dodgy knees.

(Male / 34 years old / waiting list / carpenter)

I#15: Yes, I’ve taken up golf because it doesn’t require any knee. My dad plays a bit of golf, but it’s the most frustrating game I’ve ever played in my life.

(Female / 19 years old / waiting list / medical student)

Usually the substituted and replaced activities are similar in some way – at least in terms of serving an underlying function (e.g. sporting competition). Furthermore, they were usually seen as linked by the patient themselves. A more abstract example though was provided by a patient who had just described her withdrawal from the social life of, and attendance at her church. She subsequently discussed how she had become involved in a local Patient’s Panel Group, which may well provide her with opportunities for formal engagement within a social organisation. She also saw this as a means of utilising skills and satisfying her own interests as a former nurse:

II#28: So something good has come out of it. The fact that I have been able to – although I can’t use my profession and work in my profession, I can now use it to be on these panels. So perhaps some good has come of it.

(Female / 61 years old / arthroscopy / former nurse)

Whilst the above patient describes the change in a positive light, other comments were more negative, for example, frustration with an inadequate replacement activity; embarrassment caused by a more restricted role; and upset about not being able to continue a favoured activity. One patient described a number of emotional responses including a sense of guilt when she compared herself to others more debilitated than herself, who she felt may have adapted better than she had:

I#17: Perhaps that's selfish. I don't know. It makes me feel awful because you see people in wheelchairs and that. You think, well they've adapted so why shouldn't I, but why should I adapt

...

I#17: I've adapted but I'm not happy about it.

(Female / 51 years old / arthroscopy / school support worker)

Similar to the descriptions of coping above, a number of factors were described as helping or hindering the adaptation process. The role of others was occasionally mentioned – for example, in sharing or taking over responsibility for problematic activities (such as driving) or by providing a substitute for a lost activity (see patient I#24 above). The passage of time played a part either by diminishing the memory of an event (e.g. the knee giving way) or by becoming accustomed to an altered physical state (e.g. pain or reduced mobility). An example of this was patient #21 who described how she had developed through trial and error a routine for negotiating stairs. Another patient described how the passage of time since he stopped participating in social activities meant that this altered state had just become a part of his life, even within the duration of the study (patient II#24).

Whilst many changes were enforced, some changes were at least partly the result of choice. Thus, one patient described how he balanced the desirability of playing sport and the necessity of continuing work in deciding to stop running (patient II#22). Patients described their willingness to tolerate current restriction if that reduced the risk of further physical deterioration (patient

1#17). A further example was of a woman who anticipated adaptive changes in the future:

l#35: I mean 'home', yes because obviously at some time in the future I'm either going to have to make either adjustment to my home or more. It depends on what's going to happen in the future.

MR: Mm

l#35: I may have to face the fact that some person could turn around and say, "Well, we can't replace your knee because of your additional health problems". So I may have to adjust to life and move on.

(Female / 63 years old / waiting list / retired social worker)

Loss

A clear sense of perceived loss was apparent for some patients. Perception of what had been lost usually addressed broad concepts such as a 'former self', a 'normal life' and 'general health' but also included more specific objects such as sport and sex life. Loss focused upon the function or value underlying the physical changes that had occurred in patients' lives. Thus, several patients talked about the loss of happiness, enjoyment or independence in their lives. One patient talked about a general loss of motivation, although he was also currently being treated for depression (patient l#34). The following three extracts illustrate these points:

l#24: And that's what I go through like, you know. Um, I do all my decorating, painting and whatever and I don't like relying on other people to do it for me. Um, as I say my children are as good as gold. They've been doing things for me but I just feel I'm a useless person ...

(Male / 56 years old / arthroscopy / glazier)

l#34: And, I can't really explain. Up to having that accident and that, I were happier than what I am now.

(Male / 46 years old / waiting list / water mains manual worker)

l#21: Oh! Every time something or other gives out. Social life – I haven't got none. Sex life – I haven't got none. What else is there?

Husband (l#21): I don't know girl

l#21: I haven't got a life, have I?

(Female / 49 years old / arthroscopy / housewife)

Reaction to such loss reflected a variety of emotions including desire for retrieval or restoration, anger, resentment, a sense of being maligned or persecuted, acceptance and recovery. For example:

I#24: But I just want my life back to be honest with you
(Male / 56 years old / arthroscopy / glazier)

I#21: We'd go down to the river and I'd sit on the river for hours with the binoculars watching the kingfishers and I feel that's all been taken from me ...
(Female / 49 years old / arthroscopy / housewife)

Patients commented upon some of the processes accompanying such loss. The apparent suddenness of the change and contrast with his prior state appeared to contribute to the shock felt by one patient. Some patients described choices or balances being made, for example between being physically active and suffering the painful consequences and being sedentary with a consequent loss of life quality (patient I#5). Another balance expressed was accepting some limitations or loss if other aspects of functioning could be retrieved or retained (patient II#24). Comparisons made with those around them also contributed to the sense of loss felt by some. For the patient described below two separate social comparisons are made. The first is with his friends who still are physically actively and appears to reinforce his loss. The second is with more physically disadvantaged people. Here though he finds it difficult to shift focus away from his own experience and such comparisons provides little solace for him:

I#5: I mean I couldn't imagine trying to run to the corner never alone going anywhere else, so it's a big chunk of my enjoyment taken away because I still see my friends out training and running and I still see people playing football you know – a more active social life ...
...

I#5: I think once you lose the physical side of your life, I think it drags you down mentally and in turn then you become perhaps caustic about things. You definitely become angry you know? "Why me? Why couldn't someone else have this?" I mean it's only my knees and I know there's a lot of people worse off but I don't know what they are going through, so I can only relate what I am going through and it's not a nice feeling, definitely.

(Male / 52 years old / arthroscopy / electrician)

Concomitant change

In addition to accommodation and adaptation in response to the knee problem, patients reported change in other areas of their lives which may either have impacted upon their knee or other quality of life areas.

Fundamental changes in lifestyle, (experienced or planned) were reported by some. One patient reported a process of recovery following a serious assault, several years previously which had substantially altered her life priorities (patient I#1). In contrast, another patient described her intention to change lifestyle by moving abroad (patient #30). One patient had moved since the baseline interview and was at least temporarily, pursuing a different lifestyle which indirectly served to ease her knee problems (patient II#6).

More specific changes were also reported, for example, in work. Two patients reported promotions which for one enriched this area of her life (patient II#12) whilst for the other it increased levels of stress (patient II#5). A third patient described how his increasing experience of, and confidence in, running his own carpentry business had made it easier to accommodate his knee problem:

II#22: Yes, I'm a lot more wiser being self-employed and –

...

II#22: Before I wouldn't have been confident to a customer to say, "No I don't want to do that".

MR: Right.

II#22: Now I will

(Male / 34 years old / waiting list / carpenter)

Changes in family circumstances (for example ill-health and finances) were also salient for some patients (e.g. patient II#39) and were affecting quality of

life. Changes in other, concomitant physical health problems were reported, including a general view of personal aging which appeared to affect the response to coping with the knee problem:

MR: Um, and then sport and social life are about the same as well.

*I#33: Yeah, well I'm – I'm getting older now. It doesn't matter as much now.
(Male / 38 years old / waiting list / builder)*

Discussion

Themes related to the patient's response to the knee damage and to other important events in their life have been identified which reflect different degrees of change. Amongst the themes identified above, a distinction was drawn between coping and adaptation. The former implies a more passive tolerance or acceptance of a change in one's circumstances, whereas the latter reflects a more active adjustment in behaviour. Within adaptation, different processes of modification, cessation and substitution have been delineated. Factors that may serve to facilitate or inhibit such coping and adaptive process have also been identified, for example, the use of social (or other) comparison and personality factors. Overall these response mechanisms enable individuals to accept as normal their altered state and allow new activities to replace previously desired activities. These changes are reflected in altered SEIQoL-DW cue profiles and provide a mechanism for the occurrence of re-prioritisation and re-conceptualisation response shift. Feelings of loss and the impact of concomitant changes were also evident in the assessment of change.

Longitudinal perspectives: Study patients were at various stages in the natural and clinical history of the knee problem, even after taking into account the stage of management stratification. It may be that which response process is evident may partly reflect the interval since the catalyst event. Schwartz and Sprangers speculated about whether response shift may occur simply with the passage of time.⁹⁶ Tracing response to changes in health state or maturation

has led to recommendations for longitudinal studies of both patient and non-patient samples.⁹⁴ Empirically, Wagner and colleagues noted that changes in domain conceptualisation in young people with diabetes may reflect stage of cognitive development; social conformity; and increasing maturity and life experience.²⁶¹ Whilst the first two mechanisms are unlikely to play a role in the current study, there is evidence that increased life experience is effecting change. Consideration of non-medical factors helps to illuminate this process of change.

The current evidence of coping and adaptive behaviour is consistent with other studies that have used this general approach. In a small study of young stroke patients (n=8) and their spouses (n=4), Smout and colleagues used SEIQoL-DW in a semi-structured interview.²⁶³ They found evidence of two theoretically concurrent coping strategies, assimilation and accommodation. The former was defined as coping aimed at overcoming the gap between actual and desired situation whilst the latter involved readjusting personal preferences and goals to the actual situation. Accommodation incorporated acceptance that initial goals cannot be realised.

Similarly, Richards and Folkman described how caregivers of patients with AIDS adjusted to usually negative changes.¹⁰⁹ Apparently trivial losses (e.g. changes in social life) may nevertheless have indicated important change in the couple's relationship. Grieving the loss of independence may be viewed as a necessary precursor to response shift by giving up no longer realistic expectations. Richards and Folkman applied principles of adaptive coping - individuals relinquish unrealistic beliefs about how things are and substitute new (downwardly revised) expectations consistent with revised beliefs. This enables individuals to subsequently attach positive value and meaning to the new expectations. Response shift is thus part of a necessary adaptive process, which if not achieved may cause the individual problems in coping.

The role of comparisons: Social comparison, a potential mediator or mechanism of response shift, was identified in some of the interviews.

Gibbons describes how changes in comparison habits may include a selective focus upon attributes that make the individual appear advantaged (altered self-dimensions); imagining more negative outcomes or 'worse worlds'; and manufacturing normative standards of adjustment so that one's own adjustment appears good.¹⁰¹ Self-evaluation may be a function of multiple comparisons, including one's own past and status of others on same dimension.²⁶⁴ In the current study, comparisons are evident in the section about loss. In these instances, comparisons may be maladaptive in that they serve to emphasise disability and disadvantage.

A number of factors appeared to facilitate or inhibit change. Gibbons suggests the inability to terminate a longing or desire for an irretrievable state can preclude effective response shift.¹⁰¹ This may represent the situation with some patients in the current study who, for example, were focused upon what has been lost to them. Gibbons suggest the magnitude of loss may be overestimated by recalling a lost dimension more favourably than it was. Conversely future may be viewed more pessimistically than it might turn out to be. Patient perception of the impact of the knee problem varied considerably and may reflect both personality and clinical differences. Furthermore, effective change through coping and adaptation probably requires an understanding of what has happened and what will happen regarding their knee. Uncertainty about diagnosis; natural history; and the efficacy and timeliness of clinical intervention may all serve to inhibit effective coping and adaptation – for example, for how long should a patient persist with their original expectations about knee function.

Carver and Scheier describe a model of normal and on-going adaptive self-regulation within which recalibration of goal-seeking systems and affect management systems occurs.¹⁰⁸ Although the iterative and continuous process of adjustments are usually counterbalanced and observably minimal, prolonged goal adversity may result in substantial cumulative effect. Goals may be traded for others as a form of disengagement, with primary goals being released in favour of lesser adopted goals. Nevertheless, this keeps the

individual within the same domain. In the current study, the subtle shifts found whereby different aspects of a larger domain are emphasised over time are an example of this process.

Scaling back goals helps individuals move ahead in their life. Within a hierarchy of goals of varying levels of abstraction, shift from one goal to another may serve the same underlying goal. In this way, newly adopted goals may contribute to the same core aspect. It is thus important to look beyond simple cue descriptor and to determine what is represented by any one cue (e.g. what role does 'family' play?). The level of abstraction evident in the cues provided in the study varied considerably - with some abstract concepts being provided as cues (e.g. independence), but with many more concrete examples being provided (e.g. gardening). Carver and Scheier consider re-conceptualisation and re-prioritisation as the same process and goal substitution is not indiscriminate.

8.4 Results synthesis and general discussion

This study has applied an individualised quality of life approach within an in-depth qualitative interview to concurrently assess re-prioritisation and re-conceptualisation response shift. In the three sections of this chapter:

- A case-study based analysis has provided evidence of re-prioritisation and re-conceptualisation response shift in patients with mechanical knee problems. This was found not only for patients recently operated upon (arthroscopy group), but also for those recently referred (waiting list group). The impact of re-prioritisation response shift upon summary quality of life (SEIQoL index) score was demonstrated.
- The basic specification of Sprangers and Schwartz' model of response shift has been extended by a typology of change based upon the qualitative analysis of SEIQoL data. The typology shows that reliance upon quality of life domain headers in a categorical analysis may misrepresent either real change or stability. The typology reflects adaptive

processes which may affect response shift and should inform future analysis of response shift change using SEIQoL and other measures.

- The study has provided concurrent evidence about psychological response processes that may lead to response shift, including coping and adaptation. Within the latter, several separate mechanisms have been identified (modification, cessation and substitution). Evidence of these processes further strengthens the credibility of the changes described previously, and provides a contextualised understanding of the widespread presence and nature of response shift in patients with knee problems.

General approach: This study combined individualised and qualitative methods, neither of which have been commonly applied to response shift assessment (see chapter three). However, other examples include, Lepore & Eton who used paired interviews with open-ended questions designed to assess re-prioritisation.¹⁰⁶ In other qualitative studies response shift may only have been an incidental finding (e.g. Paterson, 2004).¹²¹ In their appraisal-focused development of the original response shift model, Rapkin and Schwartz used the induction of frame of reference to consider re-conceptualisation.⁹⁵ In this model subsets of experience considered relevant to the quality of life may vary and is analogous to the changes in SEIQoL cues described in this study.

Strengths and weaknesses: Setting SEIQoL-DW administration within a broader semi-structured interview capitalises upon the natural inclination for further discussion following administration (as observed by Campbell and Whyte).²³⁶ The individualised approach highlights mechanisms of individual adaptation more than the standard needs approach and renders response shift more transparent.^{197 198} However, Macduff, notes the tension between need to reflect true change and need for reliability. Similarly, Lindblad and colleagues question whether individual methods over-emphasise real but small differences.¹⁹⁷ In the current study, interview probes helped to determine the importance of apparent change. This included direct questioning about discrepancies between recorded cues at baseline and follow-up. This approach is analogous to that recommended by Rapkin and

Schwartz's for evaluating appraisal parameters for quality of life.⁹⁵ As such this study supports the value of such questions for evaluating (confirming or otherwise) apparent changes.

Potential weaknesses of the general approach firstly include artefact in identifying response shift due to item framing, sequencing and format effects.²⁶⁵ However, consistency of data elicitation across assessments should help minimise these possibilities. Secondly, various implicit theories, and social desirability may confound the use of SEIQoL for assessing response shift.^{93 266} Thus, impression management, whereby patients either consciously or unconsciously filter information provided via self-report may affect ratings. Similarly patient expectations and denial may also affect self-report. Such biases may actually operate through any form of self-report, not simply interview-based approaches.¹⁹⁶ Whilst they may never be wholly eliminated, by extending the basic SEIQoL schedule within an interview that allowed qualitative probing helps to determine the credibility of response. Furthermore, an interviewer not aligned to a clinical service may reduce response biases related to their health or clinical care (i.e. the patient is not trying to please their clinical carer). The candour exhibited by many respondents (evident in case studies in this chapter and in appendix IX) further supports the credibility of the data.

Some specific issues raised by Joyce and colleagues about SEIQoL should be highlighted. Firstly, they describe the potential failure of respondents to nominate all cues relevant to their evaluation of overall quality of life.¹⁹⁶ The absence of certain cues which could be expected to influence quality of life (for example, family, health and employment) has been already been noted. As Smout and colleagues suggest, it may be that certain cues are possibly taken for granted by respondents or simply overlooked during cue elicitation.²⁶³ These authors suggested the use of advance notification in pre-empting respondents to reduce this possibility. Supplementing the interview with fixed prompts may address this possibility, although also increases the risk of leading the patient in their response.

Secondly, Joyce and colleagues note the complexity of deriving an overall quality of life score by identifying and weighting individual cues which are then simply aggregated. Whilst the administration process may be an imperfect representation of the constitution of quality of life, using SEIQoL as a framework for response shift assessment seems less problematic. Thus, observed changes in cue weight or ranking, supported by respondent validation may still be valid assessment of changes in the roles of individual cues. Finally, Moons and colleagues conclude that SEIQoL-DW is not a direct measure of quality of life, but rather of its determinants.⁵⁹ However, how response shift may operate to increase or decrease the influence of determinants of overall life quality makes SEIQoL particularly appropriate as a framework for assessment.

Means of observing change and stability: Observation of change in quality of life content (re-conceptualisation) and the importance of individual cues (re-prioritisation) across the interviews was facilitated in a number of ways. Firstly, following cue elicitation at baseline patients were asked whether they would have provided the same cues if they had been asked the same questions twelve months previously. Whilst the qualitative nature of the enquiry, including interviewer probing, allowed for some determination of the robustness of the patient's observation this nevertheless represents the weakest form of evidence of change as there was no contemporary assessment of cues at the 'first' time-point.

The second means of observing change was through direct comparison by the researcher of SEIQoL cues and weights between baseline and follow-up interview. The robustness of the change could be supported by contemporary descriptions of the cues by patients and descriptions of events that may have served to facilitate or hinder change. These include descriptions of physical deterioration, surgical intervention or adaptive psychological processes. A third means of observing change was asking patients at follow-up to reflect on whether they had provided consistent cues at the two assessments. This was supplemented by subsequently informing them of 'changed' cues (content or weight) and asking patients to comment upon any apparent discrepancy. This

built upon the weight of evidence provided by the second approach above, allowing the patient to confirm or contend the observation.

In addition, a fourth opportunity was provided by occasional spontaneous comments by patients regarding change or stability in cue content or weight. These were expressed in both assessments and may have been provided before cue elicitation as well at other points within the interview. Often these comments were a conscious expression of the patient's self-image. Whilst providing some insight, these observations were usually general in nature and not always consistent with change or stability observed via other means.

8.5 Summary

The combined use of individualised and qualitative methods incorporating patient case studies has identified changes in quality of life domain content and valuation. Distinct forms of change have been identified which have been used to derive a representative typology. The extent to which changes in SEIQoL cue profile represents a response shift may be informed by reference to the typology. However, simple temporal comparison of cue profiles may obscure real changes or misrepresent underlying stability. This has highlighted potential problems for the validity of group level analysis. Mechanisms that may promote response shift have been identified, including reactive coping and more proactive adaptation. Some factors that may inhibit or facilitate such mechanisms have also been identified. The concurrent assessment of re-prioritisation, re-conceptualisation and potential mechanisms serves to enhance the validity of apparent response shift change.

Preface to chapter 9

The current chapter represents two approaches to response shift assessment that have been rarely undertaken, namely qualitative interviews and using individualised methods. The analysis has addressed re-prioritisation and re-conceptualisation response shift. The following chapter continues the

response shift focus of the thesis by utilising a more commonly applied approach – the retrospective pretest-posttest design (or ‘thentest’). This approach quantifies re-calibration response shift.

Chapter 9 Evaluating scale re-calibration response shift using the retrospective pretest-posttest design

Abstract: *The following chapter assesses re-calibration response shift and its direction in the study sample. The study used a retrospective pretest-posttest design (thentest) with a common health utility instrument (EQ-5D) as the main outcome measure. The introduction to this chapter reviews the development of this study design; the implications of this form of response shift; and alternative explanations for discrepant findings revealed by the thentest. Quantitative outcome data from the patient interviews were analysed in this study. The analysis is presented within four sections addressing: (i) descriptive health outcome scores; (ii) conventional prospective change in outcome; (iii) evidence for response shift; and (iv) evidence for the validity of retrospective patient judgements. The arthroscopy patient group received a therapeutic surgical intervention following baseline assessment, whilst the waiting list group received no such intervention. The results showed that arthroscopy and waiting list patients reported similar levels of health utility and individualised quality of life at baseline, despite different likely duration of knee problems. Health utility improved over the course of the study for arthroscopy patients but not for waiting list patients. However, a response shift was found within the arthroscopy group in particular. Controlling for this re-calibration response shift resulted in no observed improvement in health utility for arthroscopy patients. Response shift was also evident in patients self-reporting deterioration through health transition items. A response shift was not found for those reporting improvement on the same items. Evidence was found which supported the validity of the thentest retrospective judgement, but there was equivocal evidence for the validity of the health transition items. Whilst the two forms of retrospective judgement appear related, they also appear to represent different judgements. In summary, re-calibration response shift was observed and appeared to act differentially according to study group. Therefore, in this study, the choice of either a prospective or retrospective approach to change measurement alters the study findings. Whilst the exploratory nature of this study suggests that conclusions should be treated*

with caution, there is some support for the validity of both forms of retrospective judgement that were assessed.

9.1 Introduction

Instrumentation bias: Campbell and Stanley identified potential sources of internal invalidity affecting a variety of study designs.⁹² One such class of variable described was 'Instrumentation' whereby measurement changes are produced by changes in the calibration of the measuring instrument, or by changes in the observers (raters). Whilst described as a possible source of bias in true experiments (i.e. pretest-posttest control group design), they felt it could be easily controlled for by use of multiple raters blind to experimental group assignment, and by random allocation of rateable material. However, as Howard and colleagues described, self-reported outcomes mean that experimental subjects are the raters, and may themselves exhibit recalibration.⁹⁰ Different experiences between study groups, even within true experiments, means that Instrumentation (re-calibration) may actually be confounded with experimental treatment.¹⁵¹ Thus, if one consequence of an intervention is the re-calibration of a patient's internal scale for response, study and control groups at follow-up may use differently calibrated internal scales for self-report. Cronbach & Furby recommended the comparison of only post-intervention scores, in randomised experiments.¹⁶⁷ However, in the light of response shift, Terborg and colleagues considered this inappropriate when self-report data was the outcome.¹⁶⁸ Therefore, other approaches to assessment have been proposed, including the retrospective pretest-posttest study design (thentest design).

The retrospective pretest-posttest design: In 1979, the retrospective pretest-posttest design was described in a series of five studies by Howard and colleagues.^{90 91} They introduced the approach as an alternative to conventional prospective pretest-posttest assessment for measuring change, specifically to address re-calibration. Their intention was to minimise the potential threat to internal validity posed by recalibration response shift when

participant self-report was used. Response shift was, therefore, primarily viewed as a bias and, as such, to be controlled for in studies.^{91 93} The threat to validity was increased when the aim of the experimental intervention was to alter understanding or awareness of the outcome being assessed, as may be the case in, for example, educational studies.⁹⁰

Subsequently, the retrospective pretest-posttest approach has become the best established and most widely used of the design methods for assessing re-calibration response shift.^{11 93 237} Early study using the approach was found in the fields of educational training and organisational development.^{89 90 156} The original studies by Howard and colleagues assessed dimensions such as dogmatism; assertiveness; and helping skills.⁹⁰ The design approach has also been applied to a number of other constructs such as attitudes towards learning difficulties; student problem-solving skills; teaching performance; interviewing skills; and communication skills.^{98 99 150 151 154} More recently, the approach has been used in a variety of clinical contexts including cancer; hearing impairment; diabetes; multiple sclerosis; stroke; and obstructive sleep apnea-hypopnea.^{94 115 117 125 129 136} Clinical outcomes assessed in such studies have included health status; health-related quality of life; health utility; and specific symptoms including fatigue.^{131 134 136 141 143 144 147} Recent empirical examples of the retrospective pretest-posttest design include assessing perceptions of safety climate in an occupational health study of construction workers, and assessment of lower urinary tract symptoms in patients with advanced prostate cancer.^{120 125}

Explanations for discrepant findings: Rapkin and Schwartz described response shift measurement as involving an account of changes that are discrepant from an expected value, hence its popularity as an explanation for paradoxical findings.⁹⁵ Schwartz and colleagues noted, and distinguished between, the use of the term 'response shift' as an ad hoc explanation for counterintuitive results, and as a phenomenon in its own right.⁹³ The abstract construct of response shift is only indirectly observable.¹⁴⁷ A consequence of the inter-relatedness of response shift with the method chosen to assess it, is that its non-detection may be either due to its true absence or to

methodological deficit (and vice versa). Therefore, the validity of the method (specifically the thentest design) continues to attract close scrutiny, and attention has also focused upon competing explanations for observed effects.

One such alternative rationalisation of results obtained using a retrospective pretest-posttest design is the application by respondents of implicit theories when formulating retrospective judgements.¹⁴⁸ Inability to recall a prior state would mean that recollection is based solely upon perception of current state and the application of heuristics.²⁶⁷ Norman, therefore, commented upon the respective validity of prospective and retrospective reports given these two competing explanations for the results from retrospective pretest-posttest design studies.¹¹⁰ Response shift theory would indicate retrospective accounts of baseline state as more valid as they are made using the same internal metric as for posttest accounts. In contrast, an implicit theories approach would suggest that prospective reports of baseline state have greater validity as they are not biased by status at follow-up.

Similar and more general concerns have been expressed by Norman and others about retrospective global assessments of change (health transition items).^{85 110} Guyatt and colleagues argued that a valid global assessment of change should be correlated with present state and show an equal correlation in the opposite direction with baseline state. However, respondents may be unable to recall their baseline state and therefore, employ implicit theories to work back from their current state.

The retrospective pretest-posttest (thentest) approach assumes that at follow-up individuals can accurately recall their baseline state and that retrospective ratings do not merely reflect recall bias. Pratt and colleagues found a dose-response effect amongst participants in an abuse-prevention programme.¹³⁷ They suggested that this supported a response shift explanation of discrepancies between baseline and retrospective scores rather than causes such as recall bias. Schwartz and colleagues found that the thentest captured both recall bias and recalibration response shift in a study of patients with multiple sclerosis.⁹⁴ The long duration of the study (over five years) may have

increased the additional noise contributed by recall bias to the retrospective rating. Visser and colleagues demonstrated convergence between a structural equation modelling approach and the *thentest* approach to the detection of re-calibration response shift.²⁶⁸ Recall bias did not invalidate the *thentest* in this study, but as the authors noted, factors such as duration between assessments may increase recall bias in other studies.

The validity of the *thentest* approach therefore remains open to question, but does nevertheless have some empirical support. For this reason the current study is largely exploratory and secondary to the main qualitative response shift studies of this thesis. Nevertheless, in conducting it, some assessment of the validity of retrospective assessments will be made.

The importance of re-calibration response shift: The original work by Howard showed that re-calibration increased the probability of the experimental hypothesis being rejected.⁹⁰ Terborg's early review of eleven studies employing both prospective and retrospective pretest-posttest comparisons concluded that in five studies, very different conclusions regarding intervention effectiveness would have been reached using the two methods. More recently, Schwartz and colleagues reviewed the clinical significance of response shift effects in 22 empirical studies (although not all were retrospective pretest-posttest designs).⁹³ They reported a median effect size of +0.25, but noted that the direction of effect was not consistent across all studies. The retrospective pretest-posttest design has thus been proposed as a means of reducing the probability of a type II error (e.g. incorrectly rejecting a 'group training program' that was actually beneficial).¹⁰⁰ However, the direction of the response shift effect may vary, and may serve to inflate, as well as reduce, the observed study effect size.⁹³

Overall aims of current analysis: The framework of the qualitative study, which incorporated quantitative assessment of health utility using the EQ-5D, permitted investigation of recalibration response shift using a retrospective pretest-posttest design (hereafter referred to as a *thentest* design). This was considered a secondary aim of the interview study, for which sample size was

derived on the requirements of the qualitative analysis. Nevertheless, an exploration of re-calibration alongside ideographic assessment of re-prioritisation and re-conceptualisation was consistent with recommendations for methodological investigation of response shift.¹¹

9.2 Methods

Sample and procedure

The sample was previously described in chapter six. A key criterion in the purposive sampling was the stage of clinical management, and contrasted patients listed at baseline for imminent interventional arthroscopy (due to be undertaken within one month of baseline assessment) and recently referred patients placed upon a waiting list. The former group (termed subsequently the 'arthroscopy' group) were expected to experience change in their physical condition. Whilst this was likely to be improvement, not all patients would necessarily benefit. The latter group of waiting list patients, who were not due to receive surgical intervention during the course of the study, were expected to remain relatively clinically stable over the study period (in practice, approximately seven months). Given long waits for arthroscopy locally, the arthroscopy group may have lived with their knee problem for much longer than the waiting list (recently referred) patients. Health utility was assessed prospectively in both interviews using EQ-5D. Each time this was the last formal component of the battery of interview assessments and was self-completed. Immediately following completion of the prospective utility assessment in the follow-up interview, respondents were asked to provide a retrospective judgement of their baseline state (thentest). In doing so, they were prompted to think back to the first interview and indicate their health as it was then.

Measures

EQ-5D (EuroQol): Valid and reliable methods for using SEIQoL to assess scale re-calibration have not yet been developed.¹³⁵ Furthermore, the time to complete the instrument twice in a single interview, and its relative complexity

precluded its use to assess recalibration. Therefore, the shorter EQ-5D was used. EQ-5D was originally developed as a standardised non-disease specific measure for describing and valuing health states.⁴The measure classifies health status across five dimensions: mobility; self-care; usual activity; pain/discomfort; and anxiety/depression. Options for each item offer three levels of response, resulting in 243 unique health states which form the basis of a utility index score.

EQ-5D is one of the most widely evaluated utility measures in health sciences research.²¹²Test-retest reliability coefficients (ICC) of 0.70, 0.78 and 0.73 (at one week, two weeks and three months respectively) have been reported for the index score in patients with osteoarthritis and rheumatoid arthritis.²⁶⁹^{#370}Evidence of construct validity has been demonstrated, for example, in a study by Hurst and colleagues with modest correlations between index score and measures of impairment, and higher correlations with subjective perceptions of disability.²⁷⁰ The measure has been shown to be responsive in patients with osteoarthritis of the knee assessed before, and six months after total knee arthroplasty.⁷⁴ A comparative review of EQ-5D and two other preference-based measures concluded that the measure showed adequate reliability and validity.²⁷¹

Transition items: In the follow-up interview, patients completed three specific transition items in which they reported firstly, how their knee was affecting their daily activities; secondly, how much time they spent worrying about their knee and thirdly, how their knee was in general (in each case, compared to six months previously). The first two items provided a seven-point adjectival response scale, and the third item a 15-point adjectival response scale. In analysis, scores for the two specific transition items were reversed so that high scores indicated improvement, in line with higher scores on the general item.

Change scores: Conventional assessment of change requires the calculation of a difference score using a baseline assessment (conventional pretest) and follow-up assessment (posttest). This value was calculated and labelled

prospective change. In addition, the difference between posttest score and thentest scores was calculated. In the context of response shift theory, Sprangers and colleagues consider this difference to represent an unconfounded measure of change and is labelled *retrospective change*.¹³⁹ Similarly, the difference between pretest and thentest scores was calculated as a measure of *response shift* (box 9.2.1). In the text, the terms 'pretest' and 'posttest' are used to refer to prospectively recorded scores at baseline and follow-up respectively (for both EQ-5D and SEIQoL-DW). The term 'thentest' is used to refer to the retrospective assessment of baseline state.

Box 9.2.1 Calculated change variables

Prospective change = posttest - pretest

Retrospective change = posttest - thentest

Response shift = pretest - thentest

Effect sizes: Prospective and retrospective EQ-5D (and prospective SEIQoL) changes were summarised using standardised effect sizes (SES) and standardised response means (SRM). SES was calculated using mean change score divided by the standard deviation of baseline scores. Only cases with valid change scores were included in the calculation of the baseline standard deviation.

Analysis plan

Four main study aims, and specific objectives (and, where appropriate, predictions) were identified:

(i) To summarise descriptively self-reported health utility and individualised quality of life

- What are the health utility and quality of life outcome measure scores at baseline and follow-up, and how do the two study groups compare?
- What is relationship between the two outcome measure variables (in cross-section)? *(expected to be moderately correlated)*

(ii) To describe conventional (prospectively recorded) change in health utility and individualised quality of life

- Is there evidence of change over time using conventional prospective assessment?

Do the two study groups differ? *(Arthroscopy group expected to show greater change – mostly improvement)*

(iii) To identify evidence of re-calibration response shift and evidence of its correlates

- Does the method of identifying change (prospective or retrospective change) affect the change detected (i.e. is there evidence of a response shift)?
- Is there further evidence of re-calibration response shift?
A response shift score was calculated as the difference between prospective and retrospective ratings of baseline state. A one sample t-test determined whether the mean response shift score was significantly different from zero.
Patients were categorised into those who reported improvement or deterioration since the baseline interview, using responses to each of the three transition items.¹³⁹ For each group, a one sample t-test assessed whether response shift was significantly different from zero.⁹⁴

139

Do the two study groups differ in apparent response shift?

- What are the correlates of observed response shift?

Potential correlates of response shift were explored: these included demographic (age, gender), clinical (study group), personality (dispositional optimism) and self-reported health transition variables. Analysis used Spearman correlation coefficients for continuous variables and t-test for group comparisons.

(iv) To identify evidence for the validity of retrospective patient judgements

- Is there evidence for the validity of the retrospective judgement of baseline state (thentest)?

The correlation between thentest score and prospective report (pretest and posttest) was assessed. (A high correlation between thentest and posttest would support an implicit theory explanation for the retrospective rating. A moderate correlation between thentest and pretest would be consistent with adequate recall of baseline state and with a response shift explanation of the retrospective rating).

- Is there evidence for the validity of self-reported health transition?

The correlations between transition items and, contemporary (posttest) and baseline (pretest) states were assessed. (Given equal variances in posttest and pretest scores, Guyatt and colleagues and Norman suggested that a valid transition item would have a high positive correlation with the former and an equally negative correlation with the latter).^{27 85}

- Are these two forms of self-report equivalent processes?

The correlations between transition items and thentest item; and between transition items and change variables (prospective and retrospective) were assessed.

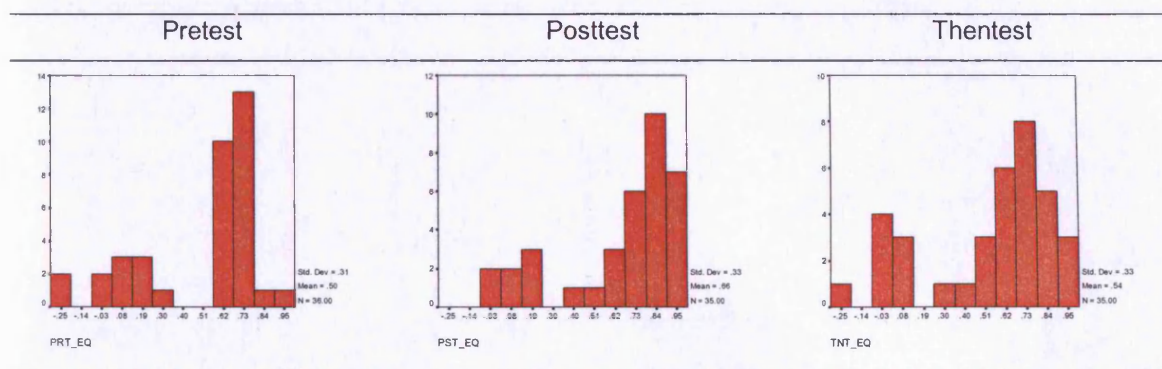
The following results section is presented within the sections indicated above. Within each section a box provides a brief summary and commentary of the findings from that section.

9.3 Results

(i) Descriptive summary of self-reported health utility and individualised quality of life

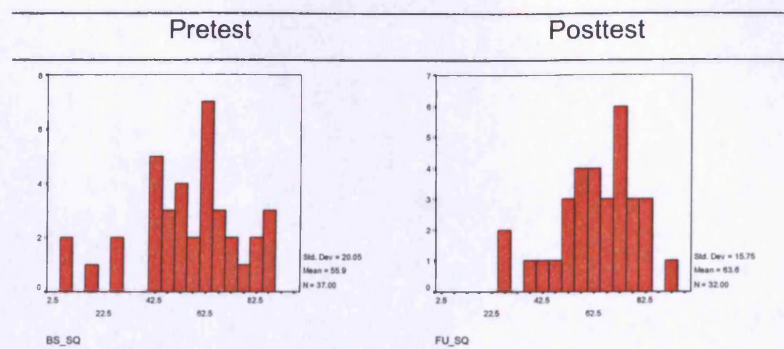
The distributions of EQ-5D scores at baseline and follow-up were all negatively skewed (figure 9.3.1). In contrast, the distributions of SEIQoL scores were less skewed (figure 9.3.2). Summary statistics for the whole sample are shown in table 9.3.1. The median pretest and posttest EQ-5D scores were 0.62 and 0.73 respectively. The median ‘thentest’ EQ-5D score was 0.62. The mean pretest SEIQoL score was 55.9 (sd: 20.05) whilst the posttest score was 63.6 (sd: 15.75). Summary EQ-5D and SEIQoL statistics for arthroscopy and waiting list patients are shown separately in table 9.3.2. Mean scores for EQ-5D are also plotted in figure 9.3.3. Differences between the arthroscopy and waiting list groups for each outcome measure at both baseline and follow-up assessments were tested using Mann-Whitney U-test (EQ-5D) and t-tests (SEIQoL). No significant differences were found.

Figure 9.3.1 Distribution of pretest, posttest and thentest health utility (EQ-5D) scores



Higher EQ-5D scores represent better health

Figure 9.3.2 Distribution of pretest and posttest individualised quality of life (SEIQoL) scores



Higher SEIQoL scores represent better health

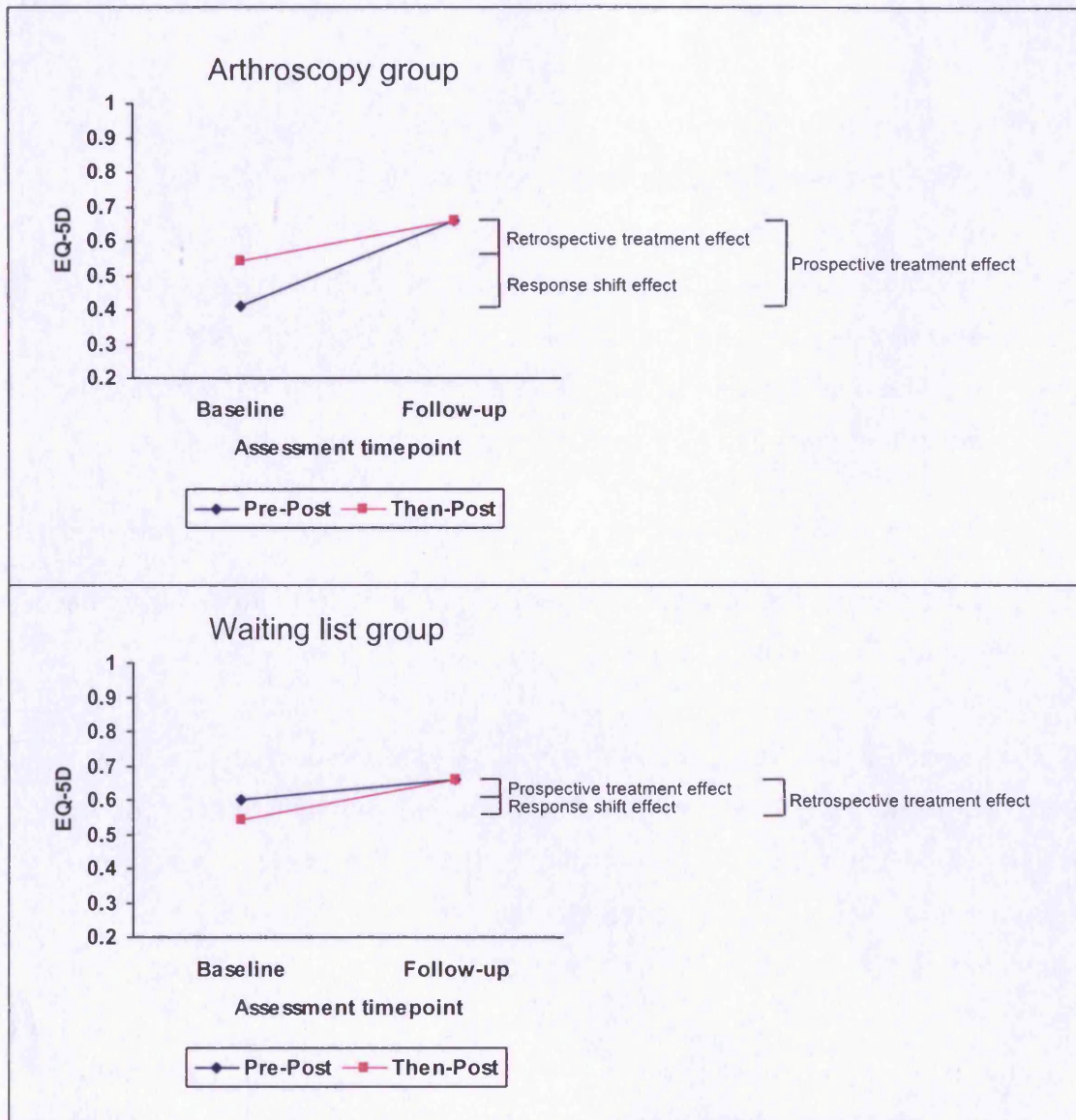
Table 9.3.1 Pretest, posttest and thentest outcome scores (all patients)

Outcome score	Median (iqr)	Mean (sd)
EQ-5D		
Pretest	0.62 (0.5)	0.50 (0.31)
Posttest	0.73 (0.36)	0.66 (0.33)
Thentest	0.62 (0.47)	0.54 (0.33)
SEIQoL		
Pretest	59.92 (24.80)	55.9 (20.05)
Posttest	64.63 (20.09)	63.6 (15.75)

Table 9.3.2 Pretest, posttest and thentest outcome scores (arthroscopy and waiting list patients)

Outcome score	Study group			
	Arthroscopy		Waiting list	
	Mean (sd)	Median (iqr)	Mean (sd)	Median (iqr)
EQ-5D				
Pretest	0.41 (0.36)	0.62 (0.60)	0.60 (0.22)	0.62 (0.10)
Posttest	0.66 (0.35)	0.73 (0.48)	0.66 (0.32)	0.77 (0.46)
Thentest	0.54 (0.36)	0.62 (0.44)	0.54 (0.31)	0.62 (0.59)
SEIQoL				
Pretest	55.25 (24.07)	60.08 (23.46)	56.63 (15.40)	53.20 (27.87)
Posttest	62.80 (15.86)	63.67 (18.98)	64.48 (16.12)	69.06 (22.55)

Figure 9.3.3 Mean EQ-5D scores at each assessment time-point



For the study sample as a whole at baseline, health utility and individualised quality of life scores were modestly correlated ($r_s = 0.53, p=0.001$).

Box 9.3.1 Commentary

- Health utility scores were negatively skewed, but individualised quality of life scores less so. Despite a longer duration of complaint at baseline, those about to have arthroscopy were no more likely to report poor health utility or individualised quality of life than those only recently referred. As expected there was a modest correlation at baseline between EQ-5D and SEIQoL, and suggests that health utility and individualised quality of life are related, but distinct constructs.

(ii) Conventionally recorded (prospective) change in health utility and individualised quality of life

Whilst baseline and follow-up EQ-5D scores were skewed, change scores were assessed using the Kolmogorov-Smirnov one-sample test and found not to differ significantly from the normal distribution.²⁷² Prospective change was therefore assessed using the paired t-test (table 9.3.3). Overall, there was a significant mean change in EQ-5D scores of 0.16 (sd=0.29, t=3.15, p=0.004).

Table 9.3.3 Prospective change in outcome scores: for whole sample, and by study group

Outcome	Group	Prospective change		
		Mean (sd)	t	p
EQ-5D	All	0.16 (0.29)	3.15	0.004**
	Arthroscopy	0.26 (0.33)	3.29	0.004**
	Waiting list	0.05 (0.20)	0.85	0.41
SEIQoL	All	4.62 (12.92)	2.02	0.05
	Arthroscopy	2.13 (10.66)	0.83	0.42
	Waiting list	7.45 (14.96)	1.93	0.07

* p<0.05, **p<0.01

When the two study groups were assessed separately, significant change (improvement) was only evident in the arthroscopy group. The mean change

for arthroscopy patients was 0.26 (sd = 0.33).⁴⁴ Assessing difference scores using the Wilcoxon signed rank test showed a similar pattern of significant and non-significant changes.

Change in SEIQoL scores was also assessed. For the sample as a whole scores improved (mean change = 4.62, sd = 12.92), although this was only of borderline significance when assessed using the t-test ($p=0.05$). When assessed separately, there was no change in scores for the arthroscopy group (mean change = 2.13, sd = 10.66) and apparent improvement in the waiting list group, although this was not quite significant at the 5% level (mean change = 7.45, sd = 14.96, $t=1.93$, $p=0.07$).

Box 9.3.2 Commentary

- As expected, arthroscopy patients reported improvement in health utility over the course of the study, whereas the waiting list patients did not. In contrast, waiting list but not arthroscopy patients appeared to show some improvement in individualised quality of life. This may reflect the distinction between these two measured constructs.

(iii) Evidence of re-calibration response shift and evidence of its correlates

For the sample as a whole, retrospectively recorded change was of only borderline significance (table 9.3.4; mean = 0.12; $t=2.07$; $p=0.046$). The standardised effect size for retrospective change was 0.35, compared to 0.50 for prospective change. The mean retrospective change score for waiting list patients was not significantly different from zero, similar to prospective change. However, for arthroscopy patients change assessed retrospectively was also not significant, unlike that found using prospective change.

Standardised effect sizes for both arthroscopy and waiting list groups were similar when derived using retrospective report (0.32 and 0.38 respectively).

Table 9.3.4 Prospective and retrospective change in EQ-5D compared: for whole sample, and by study group

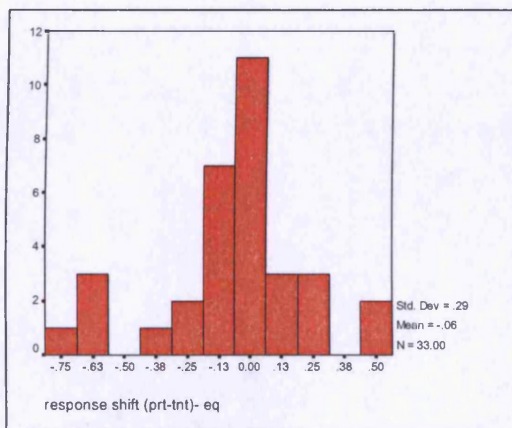
Group	Prospective change				Retrospective change			
	Mean (sd)	SES	t	p	Mean (sd)	SES	t	p
All	0.16 (0.29)	0.50 ^a	3.15	0.004**	0.12 (0.34)	0.35 ^a	2.07	0.046*
Arthroscopy ^b	0.26 (0.33)	0.68	3.29	0.004**	0.16 (0.32)	0.32	1.59	0.13
Waiting list	0.05 (0.20)	0.22	0.85	0.41	0.12 (0.37)	0.38	1.30	0.21

* $p < 0.05$, ** $p < 0.01$

a For the whole sample, standardised response means were also calculated and were 0.55 and 0.35 for prospective and retrospective EQ-5D change respectively.

The difference between retrospective and prospective rating of baseline health utility (and hence difference in changes scores) represents recalibration response shift. The mean response shift score was -0.063 (sd=0.29) with values ranging from -0.76 to 0.56 (figure 9.3.4). A positive score represented a retrospective rating of health worse than it had been originally rated.

Figure 9.3.4 Distribution of response shift scores



The goodness-of-fit of response shift scores to a normal distribution was tested using the Kolmogorov-Smirnov one-sample test and was not found to be significantly different ($z=1.035$). The arthroscopy and waiting list study groups were assessed separately for response shift using a one-sample t-

test. For the former, the mean response shift score was -0.13 ($t=-1.995$, $df=18$, $p=0.061$) whilst for the latter it was 0.028 ($t=0.371$, $df=13$, $p=0.716$).

A consistent appraisal of a subjective construct, such as health utility, should be reflected in a high correlation in scores between assessments, and also consistency of correlation with other measures. Therefore, the correlation between health utility and individualised quality of life was assessed at follow-up (table 9.3.5). Although there was a significant correlation between the two ($r_s = 0.38$, $p<0.05$), this was smaller than that found at baseline ($r_s = 0.53$, $p=0.001$). The correlation between baseline and follow-up scores was assessed for each outcome measure, and was found to be small for EQ-5D ($r_s = 0.39$, $p<0.05$) but larger for SEIQoL ($r_s = 0.59$, $p<0.001$). A similar pattern of correlations between baseline and follow-up scores was also found when each study group was assessed separately. In this case, significant correlations were only observed for SEIQoL scores (although this could also reflect the smaller sample size).

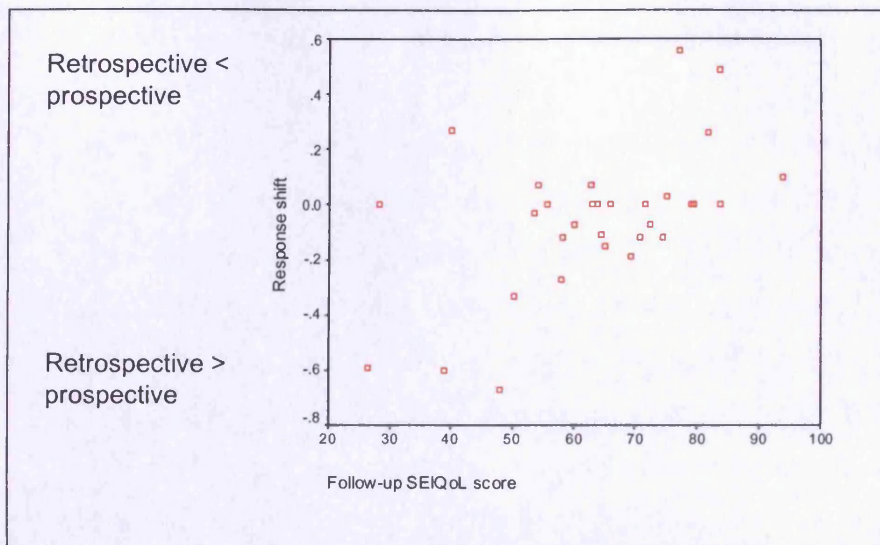
Table 9.3.5 Correlation between pre- and post EQ-5D scores and pre- and post SEIQoL scores (Spearman rho)

		EQ-5D	SEIQoL-DW	
		Post	Pre	Post
EQ-5D	Pre	0.39 ($p<0.05$)	0.53 ($p=0.001$)	-
	Post	-	-	0.38 ($p<0.05$)
SEIQoL	Pre	-	-	0.59 ($p<0.001$)

Factors associated with response shift

For the study sample as a whole, response shift was not associated with patient age; gender; or study group. It was also not associated with dispositional optimism (LOT-R). It was however, positively correlated with posttest SEIQoL score ($r_s = 0.47$, $p<0.01$) but not baseline SEIQoL score (figure 9.3.5). Higher SEIQoL scores at follow-up were associated with a worse retrospective rating of baseline utility than actually provided at baseline. Response shift score was not correlated with posttest EQ-5D score.

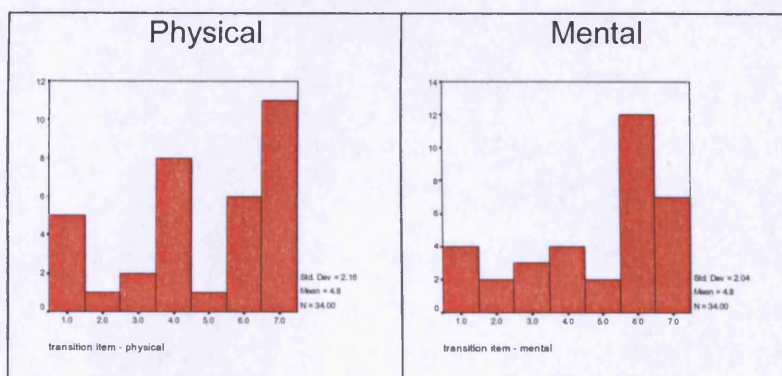
Figure 9.3.5 Scatter plot of posttest SEIQoL scores and response shift



Response shift and self-reported health transition

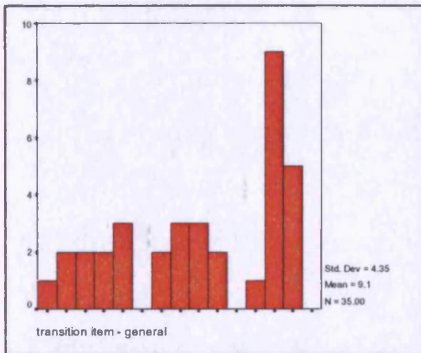
Median transition scores for physical and mental knee-related health were 5.5 and 6.0 respectively, indicating overall improvement for the whole sample (figure 9.3.6). Similarly, median scores on the 15-point general transition item were 9.0 for the whole sample, again indicating slight overall improvement (figure 9.3.7). There was no difference between arthroscopy and waiting list patients in perceived change on any transition item when assessed using Mann Whitney U-test.

Figure 9.3.6 Distribution of physical and mental health transition scores



Higher transition item scores represent improvement

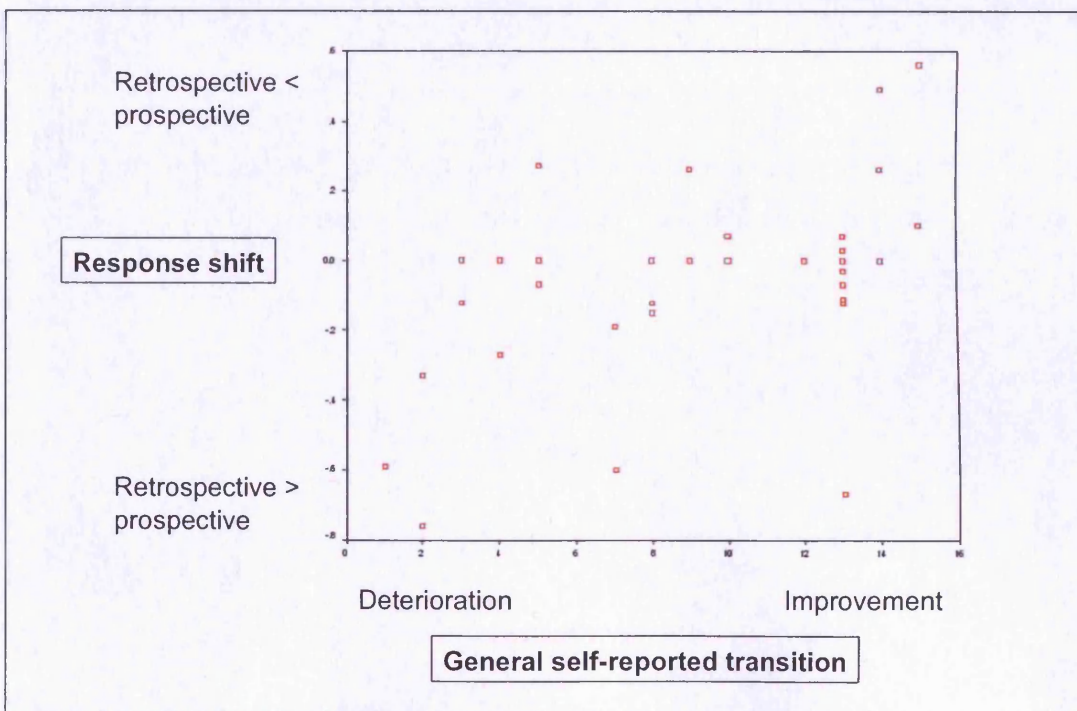
Figure 9.3.7 Distribution of general health transition scores



Higher transition item scores represent improvement

Response shift scores were significantly correlated with completion of the health transition items. Greater levels of self-reported physical or mental improvement were associated with a lower retrospective rating of health than actually provided at baseline ($r_s = -0.40, p < 0.05$; $r_s = -0.43, p < 0.05$ respectively). Accordingly, those reporting deterioration on a transition item were more likely to provide a higher retrospective rating of baseline health, than provided at the time. A similar pattern of association was found when using the general transition item ($r_s = -0.53, p < 0.01$, figure 9.3.8).

Figure 9.3.8 Plot of response shift and general transition item



The mean difference between pre-test and thentest EQ-5D scores (response shift) was tested using a one sample t-test. This was conducted separately for those reporting improvement in their condition and those reporting deterioration. Therefore, responses to the transition items were grouped according to whether the patient reported improvement, deterioration or stability. For the physical and mental transition items, 'stability' only included those responding 'about the same'. For the overall item, 'stability' was taken as those responding 'no change'. Response shift scores for each compared group are shown in table 9.3.6.

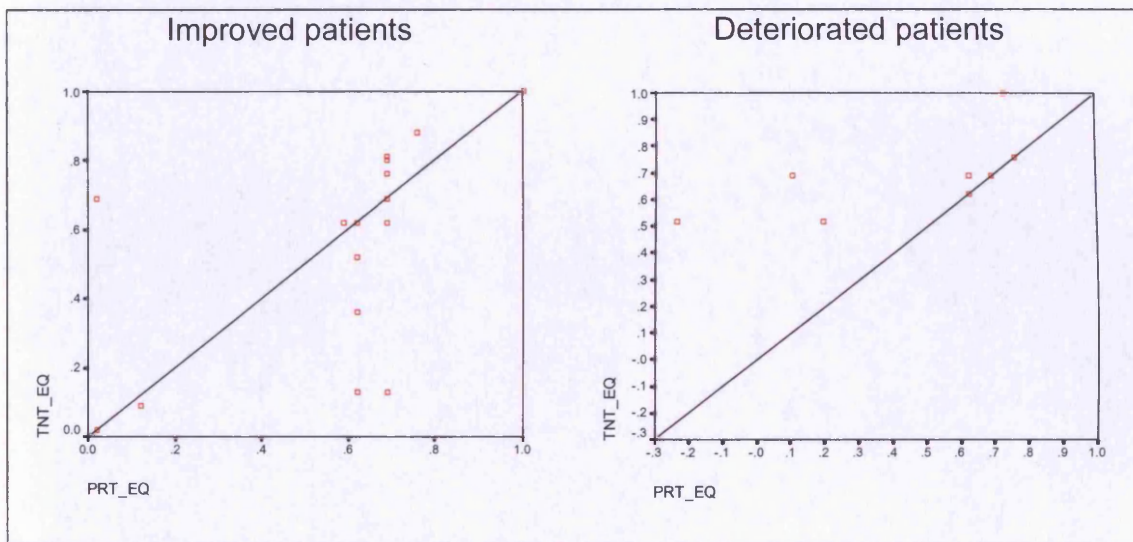
Table 9.3.6 Response shift (mean difference between baseline and thentest EQ-5D score) for improved and deteriorated patients

Self-reported change	Transition item	n	Mean difference (SD)	T	Sig
Improvement	Physical	(n=18)	0.03 (0.27)	0.42	0.680
	Mental	(n=21)	0.05 (0.26)	0.84	0.411
	Overall	(n=20)	0.05 (0.26)	0.76	0.457
Deterioration	Physical	(n=8)	-0.25 (0.29)	-2.44	p<0.05
	Mental	(n=9)	-0.24 (0.23)	-3.15	p<0.05
	Overall	(n=12)	-0.22 (0.30)	-2.54	p<0.05

There was significant response shift in those reporting deterioration on each transition item. There was no such effect shown in those reporting improvement.

To illustrate this relationship, the retrospective rating of baseline (thentest) scores and prospective rating of baseline state (pretest) scores are plotted for patients reporting improvement or deterioration on the physical health transition item (figure 9.3.9). Points along the diagonal axis indicate patients whose pretest and thentest rating were equivalent. Points above the line represents patients whose retrospective rating of baseline health was higher than actually obtained at baseline.

Figure 9.3.9 Scatter plot of actual versus retrospective (thentest) rating of baseline health: 'improved' and 'deteriorated' using physical transition item



The effect of adjusting for response shift on observed change for patients self-reporting deterioration is shown in table 9.3.7. Prospective change scores are not significantly different from zero for patients identified as deteriorated on any transition item. However, when the retrospective change score is used to account for response shift, change is now of borderline significance.

Table 9.3.7 Prospective and retrospective change in EQ-5D for respondents reporting deterioration on transition items

Change score	Transition item	n	Mean difference (SD)	T	Sig
<i>Prospective (Posttest-pretest)</i>	<i>Physical</i>	8	0.01 (0.24)	0.16	P=0.88
	<i>Mental</i>	9	0.05 (0.28)	0.58	P=0.58
	<i>Overall</i>	12	0.06 (0.28)	0.76	P=0.46
<i>Retrospective (Posttest-thentest)</i>	<i>Physical</i>	8	-0.24 (0.28)	-2.44	P=0.045*
	<i>Mental</i>	9	-0.19 (0.27)	-2.08	P=0.07
	<i>Overall</i>	12	-0.16 (0.25)	-2.20	P=0.05

* $p < 0.05$

Box 9.3.3 Commentary

- Prospective and retrospective assessment resulted in different change scores, with greater change evident using prospectively recorded observations. This difference implies re-calibration response shift. Furthermore, the effect differed between study groups with retrospective assessment increasing mean change scores for waiting list patients but decreasing it for arthroscopy patients. The observed improvement for arthroscopy patients using conventional prospective measurement was not found using the retrospective method (i.e. by accounting for re-calibration response shift, there is no improvement in health utility for arthroscopy patients). A derived response shift score further indicated a difference between study groups, although it was of borderline statistical significance for arthroscopy patients (and not significant for waiting list patients).
- The change in level of association between health utility and individualised quality of life between baseline and follow-up suggests a change in the way that one or other construct is being appraised by respondents. The small correlation between baseline and follow-up EQ-5D (compared to SEIQoL) may thus indicate a response shift change in this measure.
- There was no association found between response shift score and available demographic or clinical variables. Furthermore, there was no statistically significant difference found between the two study groups. Whilst individualised quality of life scores at follow-up were associated with response shift, there is perhaps unlikely to be a causal relationship between the two variables. It is perhaps more likely that an underlying mechanism may be influencing both variables. The nature of this relationship, therefore, requires further investigation.

- Response shift was associated with patient self-report of health change (transition). At follow-up, patients reporting deterioration in their condition were more likely to have reported a higher retrospective rating of their baseline state than they had done at the time of the baseline assessment. However, for patients reporting improvement there was no such discrepancy. Using prospective change, no difference between baseline and follow-up scores were found for patients self-reporting deterioration. However, accounting for response shift resulted in change scores of borderline statistical significance.
- These findings add support to the possible differential presence of response shift across study groups, which may therefore act as a confounding variable in experimental designs.

(iv) Evidence for the validity of retrospective patient judgements

Retrospective judgement of baseline state

The retrospective judgements of baseline health utility (thentest EQ-5D scores) were significantly correlated with both pretest and posttest EQ-5D scores (table 9.3.8). However, the retrospective judgement was more highly correlated with the baseline rather than the follow-up rating ($r_s = 0.694$ and $r_s = 0.395$ respectively). Thentest EQ-5D scores were not significantly correlated with either pretest or posttest SEIQoL-DW scores.

Table 9.3.8 Correlation between thentest EQ-5D and pretest EQ-5D, and pre- and posttest SEIQoL scores (Spearman rho)

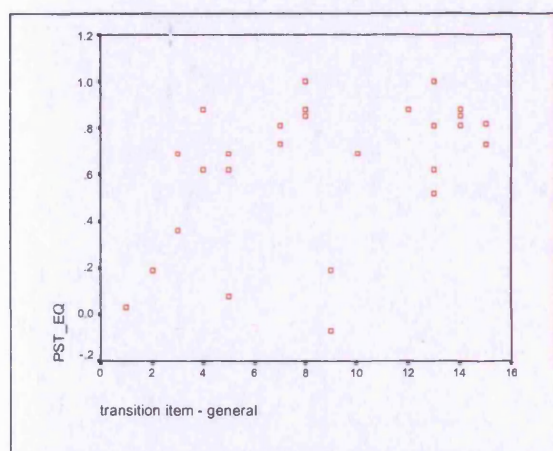
EQ-5D	Then	EQ-5D		SEIQoL-DW	
		Pre	Post	Pre	Post
		0.694 (p<0.001**)	0.395 (p=0.019*)	0.239 (p=0.173)	-0.001 (p=0.994)

Self-reported health transition

The scatter plot in figure 9.3.9 shows the relationship between the general health transition item and posttest health utility score (EQ-5D). Each transition item (physical, mental and general) was highly positively correlated with

follow-up EQ-5D score – self-reported improvement being associated with better health score (table 9.3.8). In contrast, transition items scores were not correlated with baseline EQ-5D scores.

Figure 9.3.10 Scatter plot of Posttest EQ-5D score and general health transition item



Each health transition item was positively correlated with both follow-up and baseline SEIQoL score (table 9.3.9).

Table 9.3.9 Correlation between health transitions items and quality of life scores (EQ-5D and SEIQoL)

	Transition item		
	<i>Physical</i>	<i>Mental</i>	<i>General</i>
EQ-5D rating			
<i>Baseline</i>	0.07 (NS)	0.12 (NS)	0.07 (NS)
<i>Follow-up</i>	0.61 (p<0.001)	0.55 (p<0.01)	0.51 (p<0.01)
SEIQoL rating			
<i>Baseline</i>	0.49 (p<0.01)	0.37 (p<0.05)	0.45 (p<0.01)
<i>Follow-up</i>	0.50 (p<0.01)	0.56 (p<0.01)	0.54 (p<0.01)

Relationship between the two retrospective judgements

The correlation between each of the health transition items and the retrospective judgement of baseline state (thentest) was assessed using Spearman’s rho, but no significant associations were found. This was also the case when each study group (arthroscopy and waiting list patients) were assessed separately.

The correlation between health transition items and prospective and retrospective change on EQ-5D was assessed using Spearman's rho. Significant correlations were found between prospective change and both physical and overall health transition (table 9.3.10). However, the correlation between mental health transition score and prospective change was only of borderline significance. In contrast, the correlations between retrospective change scores and each health transition item were highly significant.

Table 9.3.10 Correlation between self-reported health transition and prospective and retrospective EQ-5D change scores

Change	Physical		Transition item Mental		Overall	
	<i>Rho</i>	<i>P</i>	<i>Rho</i>	<i>P</i>	<i>Rho</i>	<i>P</i>
<i>Prospective (Posttest-pretest)</i>	0.45	p=0.01**	0.34	P=0.055	0.40	P=0.02*
<i>Retrospective (Posttest-thentest)</i>	0.77	P<0.001**	0.65	P<0.001**	0.79	P<0.001**

Box 9.3.4 Commentary

- Retrospective judgement of baseline state (thentest) was more closely associated with actual reported level at baseline than at follow-up. This does not indicate the application of an implicit theory when forming such a judgement. Rather, it indicates that patients are basing their response in part on their understanding and recall of actual baseline status. Furthermore, this retrospective judgement is not associated with contemporary assessment of individualised quality of life. Together these results provide support for the validity of the retrospective (thentest) judgement from a response shift perspective.
- The second form of retrospective judgement (health transition), was strongly associated with current (posttest) health utility score, but not with baseline state. This latter finding is contrary to what Guyatt and colleagues' state would indicate a valid transition item (that it should be equally correlated with baseline and follow-up scores but in opposing directions). However, health transition scores were moderately correlated with prospective change, (albeit even more strongly correlated with retrospective change). Together, these results provide mixed support for the validity of the transition items.
- The correlation between health transition and SEIQoL at baseline and follow-up was similar, and in the same direction. This may possibly be due to an underlying tendency for patients to both evaluate their quality of life consistently, and view change in a similarly consistent fashion. Whilst, such a conclusion would be speculative, the nature of this relationship is worthy of further investigation.
- The lack of correlation between retrospective assessment of baseline state (thentest rating) and health transition suggests that these two retrospective-based judgements are, as intended, different. However,

the greater correlation between transition items and retrospective change scores (which incorporate the thetest rating), compared to prospective change scores (which do not) suggests some degree of association. It is perhaps reasonable to expect that retrospective rating of baseline state (which incorporates patient re-calibration) and self-reported transition (a judgement of change from baseline to follow-up) would be moderately related.

- As this exploratory analysis lacks a gold standard for change, conclusions as to the validity of both change scores and transition items can only be tentative, not least because both measures are being concurrently assessed.

9.4 Discussion

Health utility and individualised quality of life were assessed in a sample of sequentially approached patients either recently referred with suspected internal derangement of the knee (waiting list group), or awaiting imminent arthroscopy for a confirmed knee problem. Arthroscopy and waiting list patients reported similar levels of health utility and individualised quality of life at baseline, even though the former group were likely to have experienced a longer duration of knee problems. Health utility scores improved over the course of the study for arthroscopy, but not waiting list, patients. However, a re-calibration response shift was found within the arthroscopy group using the thentest approach. The effect of the response shift was to increase the reported change in health utility for arthroscopy patients. Accounting for response shift, therefore, reduced the observed change in health utility. Response shift was also evident in patients who reported deterioration through health transition items. In contrast, response shift was not found for those reporting improvement on the same transition items. Evidence was found which supported the validity of the thentest retrospective judgement of baseline health utility. However, evidence for validity of the health transition items was equivocal. Whilst the two forms of retrospective judgement appear related, there is also evidence that they represent different judgements.

Mean baseline scores for arthroscopy patients were lower than for waiting list patients but the difference was not statistically significant. The mean utility score for arthroscopy patients at baseline (intended to be pre-operative, although some patients had just received their arthroscopy) was 0.41. This was somewhat lower, for example, than that found for a sample of 84 patients assessed immediately following uncomplicated arthroscopic partial meniscectomy by Goodwin and colleagues (0.54 and 0.56 for two study groups).²⁷³ Mean utility score for waiting list patients was 0.60 - similar to that reported by Hollingworth and colleagues for a sample of patients referred for

MRI of the knee (61.3).¹⁷³ Length of total waiting times for arthroscopy was not reported by Goodwin and colleagues. However, it may be that extended waits for arthroscopy for patients in this study (reported standard waits of between 3.75 and 5.75 years) has resulted in greater deterioration in health utility.

Prospective and retrospective change and response shift: As expected, when assessed prospectively, significant improvement in health utility was found for patients listed at baseline for imminent arthroscopy, but not waiting list patients. For the former, mean utility score was 0.66 at follow-up (approximately seven months). This compares to 0.75 and 0.81 in post-surgical assessment of two groups (at six weeks) recorded by Goodwin.²⁷³ However, using a retrospective assessment of change resulted in a reduction in effect size for the arthroscopy group and increase for the waiting list group. Using the retrospective measure of change, neither group exhibited significant change.

That arthroscopy patients retrospectively rated their baseline state higher (better) than they originally rated at the time of baseline assessment is indicative of a re-calibration response shift. In practice, this had the effect of reducing the level of reported change to a non-significant difference. Why would these patients in retrospect rate their pre-operative health utility higher than they did at the time? From the follow-up interviews it was clear that patients varied in the extent to which they felt they had benefited from their arthroscopy, and the extent to which they had yet recovered function. One possible explanation is that some dissatisfaction with outcome may have led patients to re-appraise their pre-operative (i.e. baseline) state in a more favourable light.

Response shift and change in individualised quality of life: In contrast to health utility scores, there was no change in individualised quality of life scores for the sample as a whole. However, there was a trend for improvement for waiting list patients, rather than for patients who received arthroscopy, (although this did not reach statistical significance). For the

former group, such an improvement in quality of life, in the absence of a medical intervention (or improvement in health utility), also suggests that some adaptation and response shift may have occurred. These data may emphasise that health status is but one contributor to individual quality of life, albeit an important one for many people. Patients may respond to disease and disability by focusing away from health and onto other aspects of their life (i.e. coping and adaptation facilitating response shift). If so, reliance upon SEIQoL index scores may result in a relatively unresponsive tool for evaluating interventions. Instead, exploring cue profiles may prove even more insightful for understanding what is of importance to individuals and how this changes over time.

Health utility and individualised quality of life were moderately correlated at baseline. Given the health-specific focus of the former measure, and the ideographic nature of content provided for the latter, this was expected. It is therefore, interesting to note the weakening of this association at follow-up. It may be that at baseline, when health status was worse, patients were more influenced by their physical condition when considering quality of life than they were at follow-up. Alternatively a re-calibration in health utility response, due to health deficit, may have changed the nature of the relationship between utility and individualised quality of life.

Correlates of response shift: The only variables associated with response shift change were follow-up SEIQoL score and health transition. Other studies which have addressed this question have also found few correlates with response shift. For example, Bar-on and colleagues found no relation between re-calibration and age and education.¹³³ Similarly Rapkin found few socio-demographic predictors of change in personal goals.¹³⁸ Manthei investigated explanatory variables (including age and gender) for differences in the presence of response shift and found no associations.¹⁴⁵ Bernhard and colleagues found no consistent pattern of modelled biomedical or socio-demographic factors affecting quality of life 'reframing' identified via the test.¹⁶⁴ Finally, Schwartz and colleagues found relatively few demographic factors associated with recalibration response shift scores in a

sample of patients with multiple sclerosis.⁹⁴ The lack of obvious socio-demographic correlates with response shift may reflect the greater role that other psychological mechanisms may play in adaptation.

The status of retrospective ratings: The thentest retrospective ratings of baseline state were correlated more highly with actual baseline scores than with follow-up scores. A response shift explanation of retrospective report suggests that accurate recall of baseline state is combined with an amended internal metric to produce a renewed judgement. An alternative approach suggests that perception of current state is combined with an implicit theory of how (in this instance) health may have changed, to form this judgement. Therefore, the results favour the former explanation for the retrospective rating, and provide support for the validity of the thentest design.

The status of transition items: Overall, self-reported health transition items indicated improvement for the sample, although a small number of patients reported deterioration. Whilst health transition items were used in this study to stratify respondents into improved and deteriorated groups, their self-reported nature means that they are also susceptible to response biases and caution has been expressed about their use as criterion variables.^{27 144 237}. Self-reported change was correlated with both health utility and individualised quality of life scores at follow-up. The strongest correlation was between the physical transition item and EQ-5D rating, and the weakest between the same transition item and SEIQoL. The former high correlation is unsurprising given the impact the knee problem was likely to have across the EQ-5D dimensions, four of which could be viewed as related to the 'daily activities' referred to in the transition item. Although similar activities may be encompassed within SEIQoL content, this outcome measure obviously allows for consideration of more diverse factors, not necessarily affected by the knee problem.

The lack of an association between transition items and baseline EQ-5D would endorse Norman's caution about the item's validity and indicate the application of an implicit theory.²⁷ Transition item scores were equally correlated with both follow-up and baseline SEIQoL scores. This doesn't

necessarily refute an implicit theory explanation, as correlations with both follow-up and baseline scores are positive. It may, for example, indicate a response style bias to be either positive or negative in outlook. The potential to express such an outlook may be greater with the ideographic SEIQoL measure than with the standardise items in EQ-5D. Such an explanation is also consistent with response shift - people with better adaptation skills may also be more likely to report periodic improvements (via transition item) and at any one time-point orientate their focus of interest to emphasise aspects which enhance quality of life.

Transition items and retrospective thettests both represent psychological construction whereby judgements are formed by re-construction of memory for prior states.¹¹⁰ Potentially moderately related, in the current study there is evidence that the two processes are somewhat independent. The greater correlation of thettest with pretest rather than posttest supports the validity of the thettest. The correlation of transition item with posttest but not baseline may raise doubts about the validity of the transition item. Nevertheless, the apparent distinction between the thettest and transition item responses also suggests that patients are not simply using perception of their current state when forming all retrospective judgements.

When the transition items were used to form separate groups of improved and deteriorated patients, a significant response shift was found for the latter group only. The size of this effect illustrates the different conclusions that may be drawn depending upon whether prospective or retrospective change was used. It also further demonstrates the potential for response shift to confound an experimental design where there is a differential therapeutic effect.

Comments about the thettest approach: Some methodological comments about the use of the thettest approach in the current study are worth noting. Including additional outcome (thetest) questions increases respondent burden and, as such, may increase the risk of problems such as reduced response rate.¹²⁰ Schwartz and colleagues addressed this burden by selecting 'best' items to represent desired subscales, and selecting 'best'

subscales to represent key domains within a quality of life dimension.⁹⁴ Within the current study, the brevity of the EQ-5D (five items) minimised additional response burden and the items were located at the very end of the interview. Whilst the full follow-up interview itself was quite lengthy, the semi-structured and interactive interview probably helped maintain interest and motivation, and perhaps more so than a self-administered questionnaire.

An optimal recall period for the thentest is not apparent and may vary with the outcome being assessed. Guyatt and colleagues assessed recall in three studies each using an interval period of four weeks.⁸⁵ They considered that the salience of clinical encounters and data collection may serve to enhance recall of baseline states. However this advantage may be reduced if many visits (assessments) are involved. Whilst the time interval in the current study was about seven months, intervals in other thentest studies have ranged from 25 minutes to five years.^{94 153} For patients to make a valid renewed judgement, sufficient time must have elapsed for an adaptive change to have occurred, but not so long that patients can not recall their original state. As patients were being asked to recall a particularly salient event (either referral to orthopaedics or arthroscopy) their recall of baseline state may be expected to be fairly good. Furthermore, the interview itself may have served to enhance recall with its focused discussion, for example, of clinical history and in doing so provided retrieval cues for respondents.¹⁴⁴ Finally, few patients expressed difficulty in recalling baseline state when completing the thentest. Although in both thentest and transition items, patients were instructed to think back six months, the actual duration since baseline was on average slightly longer. Patients were nevertheless generally encouraged to respond according to their state at the baseline interview.

Objective criteria for change: Sprangers and Hoogstraten both described how the absence of a criterion measure of change limits conclusions about the relative validity of conventional change and thentest (retrospective) change.^{101 151} Several studies have found a higher correlation between objective measures of change and thentest assessments.^{90 98 100} However, using objective measures has not always helped to clarify the situation. For

example, Ahmed found objective performance-based measures failed to distinguish patients exhibiting response shift. Nevertheless, an objective measure of change in the current study would have been useful -especially if the thentest had been the primary purpose of the research interview. For example, this could involve an independent (clinical) assessment of functional status. Whilst an objective measure would also help in future studies of the validity of retrospective judgements (such as health transition) there is always going to be some deficit in objective assessments if the target construct is essentially subjective.

Most of the quantitative analysis in this thesis is presented in the current chapter and a large number of statistical tests were used. This raises the possibility of a type I error occurring. A conservative approach to multiple significance testing could have been to adopt a higher threshold for statistical significance (for example 1%). However, in the currently presented analysis care has been taken not to place too much emphasis on absolute probability values. This is partly driven by the exploratory nature of the analysis presented and recognition of the relatively small number of cases included in the analysis. Interpretation of these study results should be therefore be viewed as suggestive rather than conclusive. Furthermore, as much as possible clear analytic objectives were specified a priori to reduce unnecessary testing.

9.5 Summary

The thentest study was conducted as a secondary component of the main qualitative response shift interview study. As such it was primarily exploratory in nature, and a formal sample size calculation was not conducted.

Nevertheless, changes were detected which suggested that re-calibration response shift occurred both within the arthroscopy and waiting list group of patients. That these changes occurred for the health utility and individualised quality of life outcomes respectively emphasises the distinction between these two outcomes. It is also a reminder that adaptive processes may result in multiple and concurrent response shift changes. The study provides evidence that supports the validity of a response shift explanation of retrospective change, but is also equivocal about the validity of transition item responses. As both processes involve psychological reconstruction in forming response, the validity of thentest and transition items continue to require close scrutiny in future study. Finally, although response shift changes were apparent, no socio-demographic correlates were found. Although this is common to other response shift studies, further work with a larger sample may be informative.

Preface to chapter 10

The presence, nature, and quantification of re-calibration, re-prioritisation and re-conceptualisation response shift have been addressed in the chapters 8 and 9. Furthermore, the mechanisms contributing to response shift have been explored. However, in the following chapter one particular process is explored in more depth – the link between the respondent's perception of cue level (how satisfied they are with each important life domain) and the respondent's weighting of that cue (how important they rate it). The analysis uses both qualitative and quantitative data from the response shift interviews.

Chapter 10 Exploring the relationship between cue level and cue weighting

“But when it comes to saying in what happiness consists, opinions differ, and the account given by the generality of mankind is not at all like that of the wise ... and often the same person actually changes his opinion: when he falls ill he says that it is health, and when he is hard up that it is money.”

Aristotle, The Nichomachean Ethics²⁷⁴

Abstract: *During the qualitative analysis of interview data presented in chapter eight, incidental evidence emerged which suggested a link between level of quality of life cue and the degree of importance attached to the cue. Therefore, secondary analysis of the relationship between satisfaction with cue level and weight was conducted, using both quantitative SEIQoL data and descriptive qualitative interview data. The exploratory correlational analysis provided support for the link between cue weight and level, although this varied with study group and timing of interview. Descriptive analysis provided evidence that the relationship between weight and level may be both positive and negative. Perception of a change in cue level, rather than simply stable high or low cue levels, may also effect change in cue weighting. This suggests a potential mechanism for response shift, whereby re-prioritisation is partly consequent upon actual level of quality of life domain. If the selection and weighting of life areas was influenced by their actual level, this may also suggest a means whereby quality of life homeostasis can be maintained.*

10.1 Introduction

From the conduct of the qualitative analysis of SEIQoL data presented in chapter eight, it appeared that respondent reports of cue level was in part related to the weighting they assigned that cue. Bayle and colleagues commented upon the need for respondents to distinguish between the importance they assign to a cue and their level of satisfaction with it.²⁵⁶ The relationship between SEIQoL cue weight and level has been addressed by

Moons and colleagues who applied several criteria in their psychometric evaluation of the SEIQoL, including assessing the internal structure of the instrument.⁵⁹ Like Bayle and colleagues, they noted the potential for respondent misunderstanding between cue level and cue weighting - which would manifest itself as a high correlation between the two scores. Therefore, they proposed a 'low to moderate' correlation between the two to be an acceptable marker of valid scale completion.

Moons and colleagues conducted a cross-sectional survey of 579 patients with congenital heart disease, and a longitudinal assessment with a subset of 130 individuals.⁵⁹ They found a correlation of 0.26 ($p < 0.001$) between cue level and weight, although it was not made clear whether this referred to the former or latter group. The study left a number of questions unanswered. For example, no additional information about this relationship was provided and details of data collection method have yet to be made available. Neither is it clear why the authors considered a low to moderate correlation (rather than no correlation) to be acceptable.

If cue level at least partly determines cue weight this could be one mechanism by which re-prioritisation response shift may occur. The availability of qualitative interview data also allowed potential for an exploration of the meaning of cue weighting by individuals. It was therefore, decided to address this issue through secondary analysis of the current study data. Therefore, the analysis aimed to:

- determine whether cue levels are independent of the weights accorded to them by respondents (evidence that the two are strongly related may suggest a mechanism for re-prioritisation)
- explore evidence for the valid completion of SEIQoL (for example, whether patients clearly distinguished between cue level and cue weight when providing their response)

10.2 Methods

The analysis sought to utilise and synthesise both the quantitative SEIQoL data, and descriptive data from the qualitative interview.

Quantitative analysis: Raw SEIQoL scores and other quantitative data from the interview had previously been entered onto SPSS for quantitative analysis (chapter 5). The analysis was conducted at cue level. Therefore, data were extracted and input into a new system file with each record representing a cue weight and cue level combination. Additional variables identified for each pair of values were also included in the file. These were unique patient identifier; study interview (baseline or follow-up) and stage of clinical management (arthroscopy listed or referred).

Correlation coefficients for the relationship between cue level and cue weight were calculated for the sample as a whole. The same relationship was also assessed separately for cues derived at baseline and follow-up; and for the two principal study groups. Finally, to determine whether there were individual differences in associations between cue level and weight (and to account for clustering within patient cue responses), rank correlation coefficients were calculated for each patient for whom ten complete pairs of scores were available. These resultant correlations were plotted using a histogram.

Qualitative analysis: As part of the modified SEIQoL assessment, patients were asked to expand upon and describe their ratings of cue level, and their reasons for weighting cues as they had. As part of the thematic analysis of the interview, instances where patients had described how they derived cue weights were identified and are analysed descriptively. Results are presented in the text with illustrative quotes.

10.3 Results

Quantitative analysis

Potentially, 390 cues would have been generated if all interviews had been followed up; SEIQoL had been conducted; and five cues nominated at each assessment. In practice, 345 cues were available for assessment. The maximum cue level reported was 100 and the lowest was 0 (mean: 57.2, SD: 28.4). The highest cue weight provided was 81 and the lowest 0 (mean: 20.0, SD: 10.6). The distribution of cue level scores was multimodal with peaks around the middle of the scale, at 80 and at 10 (figure 10.3.1). In contrast the distribution of cue weights was unimodal and positively skewed (figure 10.3.2).

Figure 10.3.1 Levels for each SEIQoL cue elicited at baseline and follow-up interview

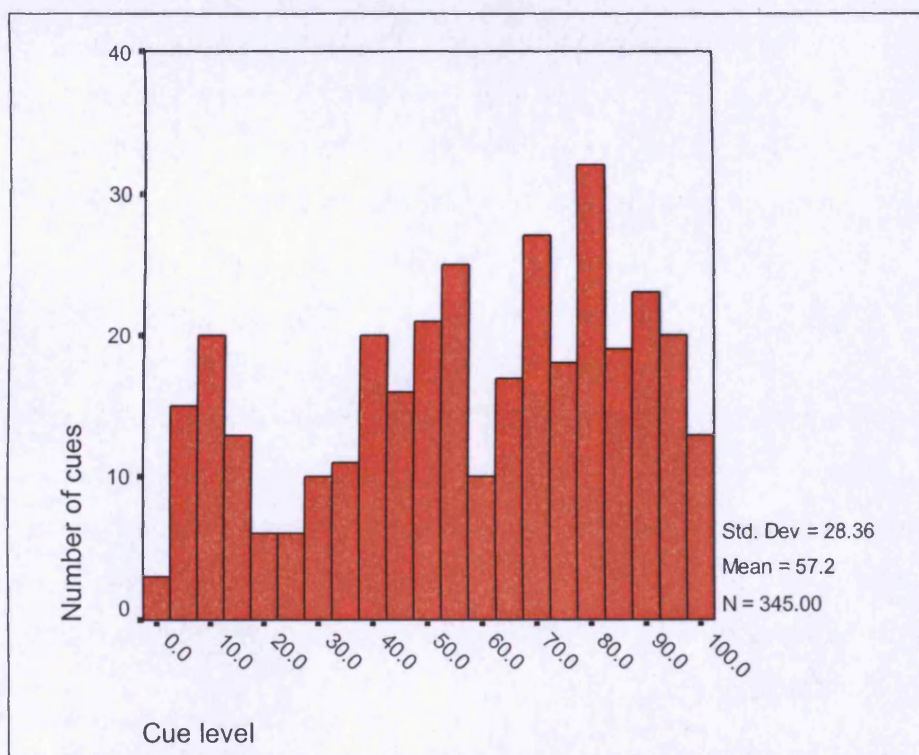
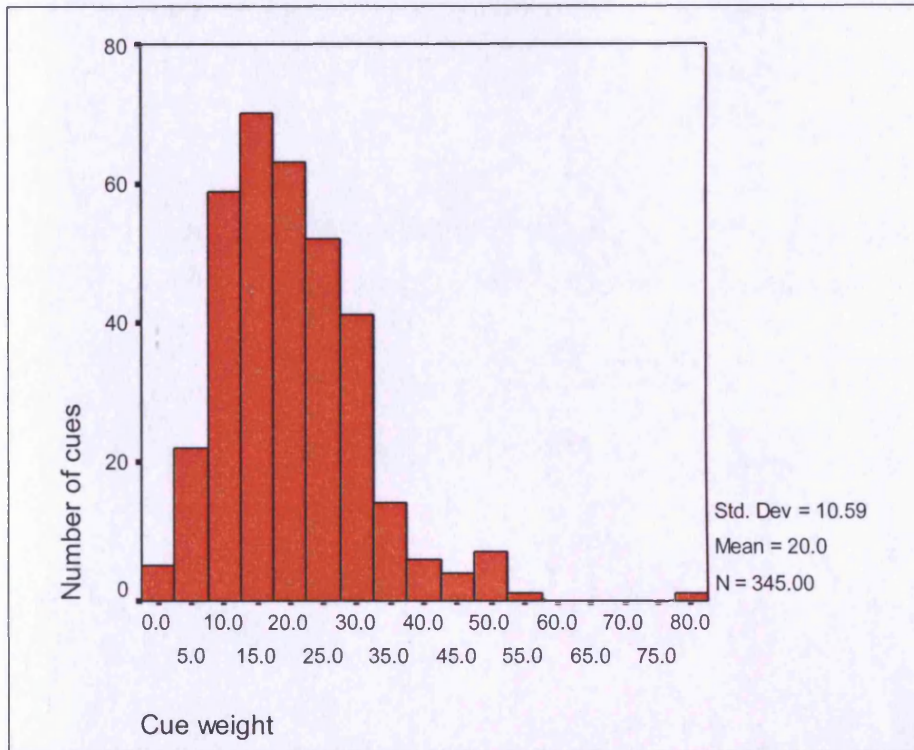


Figure 10.3.2 Weights for each SEIQoL cue elicited at baseline and follow-up interview

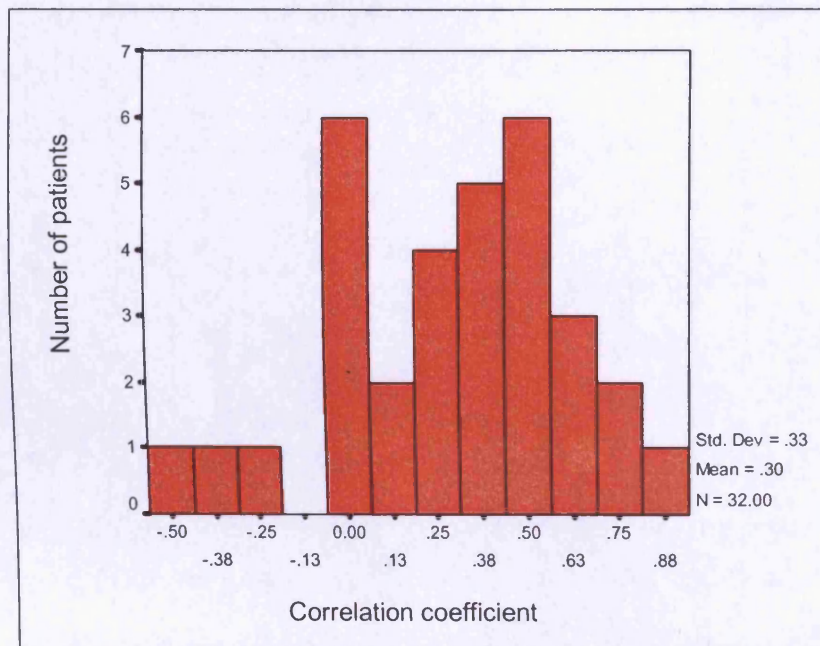


When baseline and follow-up cues were considered together there was a significant positive correlation found between cue weight and cue level ($r = 0.152, p=0.005$). However, when cues were assessed separately for each timepoint, there was no correlation found between cue level and weight at baseline ($r = 0.089, p=0.229$). The relationship between weight and level was still significant at the follow-up assessment ($r = 0.237, p=0.003$). When cues were considered separately for the two study groups there was found to be a significant relationship between level and weight for arthroscopy patients ($r = 0.259, p<0.001$). However, no such relationship was found for waiting list patients ($r = 0.054, p=0.490$).

There were 33 patients with a complete set of SEIQoL cues. The rank correlation between cue weight and cue level was calculated for each individual patient for whom ten cue points were available. A histogram of the resultant correlation coefficients is shown in figure 10.3.3 below. The bimodal distribution shows one peak around 0, implying no correlation between cue

weight and cue level, and one peak around 0.5. The latter suggests a positive correlation between weight and level. Interestingly, for three patients there was a negative correlation found. The overall pattern suggests that there may well be variation between individuals in the way that they construct their responses to these questions.

Figure 10.3.3 Rank correlation coefficients (cue weight and level) for patients with complete data at baseline and follow-up



For illustrative purposes, individual patient scatter plots were produced for patients at three different points of the histogram to show examples of a negative, positive and no correlation. These are shown in figures 10.3.4 to 10.3.6 below. The five data points from the baseline assessment are shown in red and those from follow-up in green.

Figure 10.3.4 depicts a negative correlation. It would appear that one point in particular from the baseline assessment (with the lowest cue level) exerted a particular influence. This cue was *family* and the patient's reported level resulted from difficulties caring for both an elderly parent, and a grandson with developmental difficulties. His own knee problem was adversely affecting his

family life in that he was less confident in safely caring for his grandson, and was restricted in playing with his grandchildren in general.

Figure 10.3.4 Scatterplot for patient with negative correlation between cue level and weight (patient ID #7; $\rho = -0.437$)

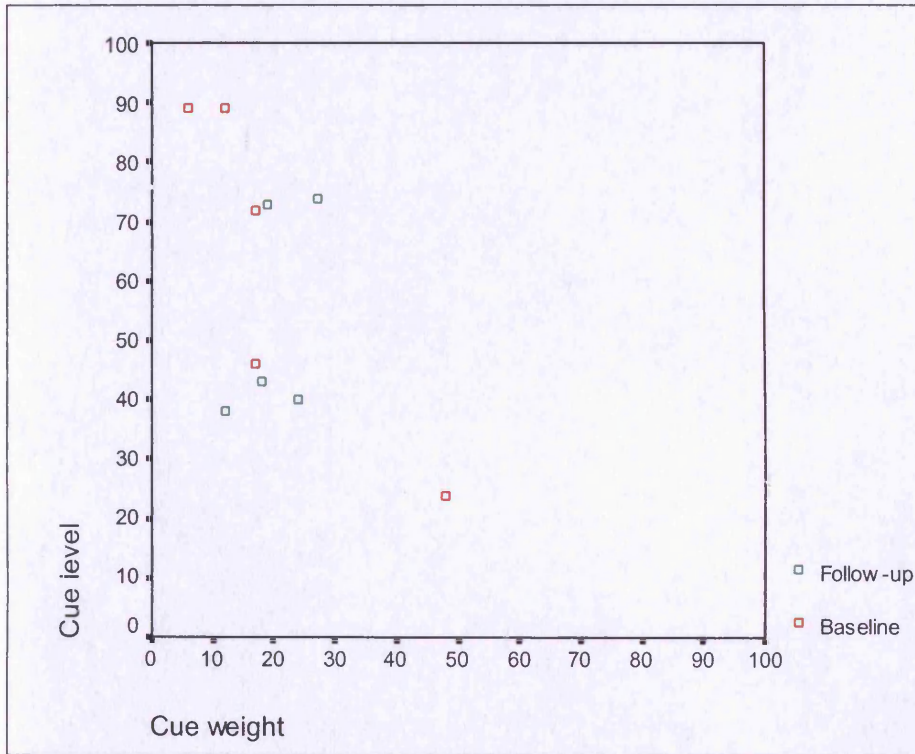
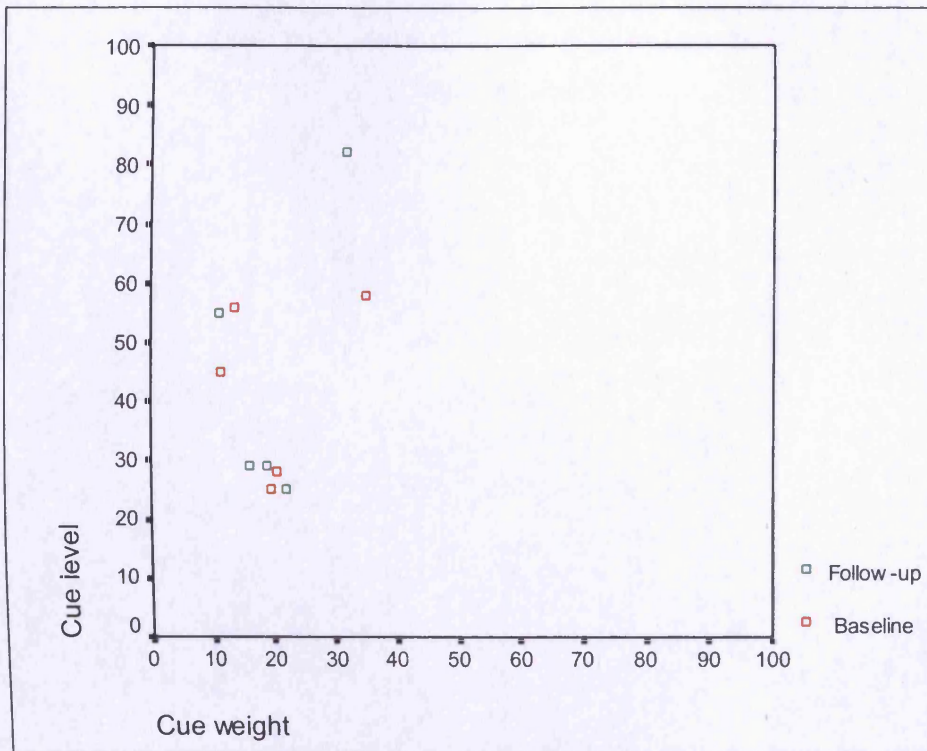


Figure 10.3.5 depicts a patient (#19) for whom there was no observed correlation between weight and level. The highest weighted cue at follow-up was *family*, which also had the highest level. This cue did not feature at all at baseline. The same was also true of the second highest rated cue at follow-up - *health*.

Figure 10.3.5 Scatterplot for patient with no correlation between cue level and weight (patient ID #19; $\rho = 0.037$)

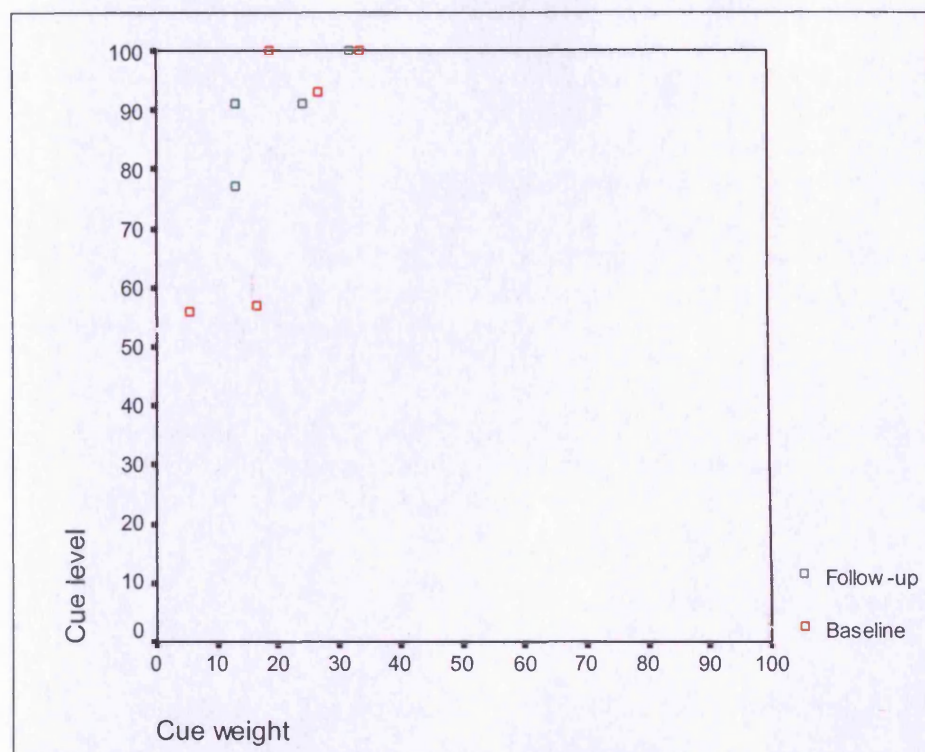


For patient #20 there was a positive correlation between cue level and weight score (figure 10.3.6). The lowest weighted cue at baseline was *work* for which was also recorded the lowest level of the five cues. Similarly, *general health* was also given a low weight and level. In describing his cues during the baseline interview, the patient commented upon how the weight he attached to work was reduced because he was unable to return to employment following his injury:

*I#20: Work at the moment is down at the bottom of the scale because I can't a
 – do any work -*

(Male / 29 years old / listed / security guard)

Figure 10.3.6 Scatterplot for patient with positive correlation between cue level and weight (patient ID #20; $\rho = 0.754$)



Note: Two cues intersect at weight 18.5, level 100, one at baseline and one at follow-up

Descriptive qualitative analysis of association between cue level and weight

Many respondents overtly linked the importance of an individual quality of life domain with their self-reported status (level) for that domain. When this occurred, a causal relationship was generally suggested, usually with level of functioning determining the consequent weighting. Patients described situations where poor status resulted in either an increased or decreased weight being attributed to the cue. Similarly, a good status was also found to be related to both increased and decreased weights. Examples of each type of relationship are described in the extracts below:

a) Lower status and higher weight:

Patient #35 described how previously she would have accorded greater weight to family matters due to the needs of her, now deceased, stepfather.

Patient #38 described how difficulties at work would have caused her to attach more importance to it:

MR: And do you think that you would have put them into this kind of order with health first, and then family and home, or do you think that it might have been a different ...?

I#35: A year ago there was different perspective.

MR: Right.

I#35: My stepfather was alive a year ago and was causing a great deal of problems within the house, a great deal of family dysfunction.

(Female / 63 years old / referred / retired social worker)

I#38: Um, work might be slightly higher [than her current rating] because I wasn't happy in the school I was in and things were changing there.

(Female / 47 years old / referred / teacher)

b) Lower status and lower weight:

In contrast to the patient above, patient #14 below reacted to work problems by describing a reduced weighting for work. Patient #15 had been prevented from pursuing sporting activities and therefore, recognised other areas of life as currently more important:

II#14: ... the one that I did contemplate a bit was the career one because that is a reasonably important part but, at the moment, because things aren't going that well I didn't feel that it was that important.

(Male / 37 years old / listed / civil engineer)

II#15: And because I can't do sport as much anymore, the other two are going to be above sport at the moment.

(Female / 19 years old / referred / student)

c) Higher status and higher weight

The following patient had recently spent much more time with his family and had developed an increased appreciation of their value:

II#19: ... and it was just good to be with them ... when you kind of get back together with them and see and stuff that you realise you are missing something you just didn't realise the rest of the time ...
(Male / 24 years old / listed / medical student)

d) Higher status and lower weight

In response to the initial cue elicitation question, patient #9 discussed that he had little work-related pressure and subsequently did not provide work as a cue. He also described how over a longer period of time the importance of money had decreased as he had become more financially secure. Patient #31 described how the importance of health had declined as his immediate health concerns had receded:

II#9: I own a company so I've got no pressure of work, I can come and go as I please ... so I've got no pressures there.

II#9: No I suppose when I was in my twenties money would have been a much bigger factor because I hadn't got any. So yeah, as you get more behind you it becomes less of a factor.
(Male / 51 years old / listed / businessman)

II#31: Well the thing is I'm in good health now, you know, it's less important if you know what I mean.
(Male / 57 years old / referred / newspaper editor)

Cue selection and weight: As can be seen from several of these extracts, respondents consciously linked their assessment of cue importance and their status. If the level of one domain of life (and therefore, a potential SEIQoL cue) contributed to its weighting, then this could be expected to influence its initial selection and nomination by the patient. This was most obvious where a cue had changed, either by being newly introduced or removed from a profile. Patient #26 provided an example of this when she described her choice of a new cue (*moving house*) due to a recent deterioration in that regard:

MR: Do you think you would have come up with the same five areas that you mentioned today?

I#26: No. I would have with my sons, yes. The moving – no, because it wasn't that bad round here ...

(Female / 51 years old / referred / not working)

In contrast, patient #18 had nominated at baseline, the cue *cycling*, with which she was experiencing difficulties due to her knee. Subsequently she did not nominate it, despite resuming the activity (although she had provided a more general cue of *mobility*). She commented:

II#18: ... I think I ride it more now than when you came before because I was having problems ...

(Female / 64 years old / referred / retired)

Similarly, the following patient reported health as a cue for the first time at follow-up, as her knee problem had become more debilitating:

MR: Do you think there's any reason why perhaps you've mentioned it [health] now as one of your five important areas?

II#38: Well I'm sure there is. You know, I am finding it more difficult to get on with things.

(Female / 47 years old / referred / teacher)

The final patient also described why *physical activities* was only nominated as a cue at baseline, a time when such activities were particularly restricted by her knee problem:

II#6: I think maybe because I was at home and my knee was bad when I couldn't do any physical activity it came up as more important, but now I'm down here – do just generally more physical activities ...

(Female / 19 years old / referred / call centre operator)

Changes in cue level: Whilst patients appeared to attribute importance to either high or low cue status, it was also apparent that a comparison was being made between current and previous (or expected) level. It is possible therefore, that the actual *change* in level led to re-prioritisation. The change in cue level may have been relatively recent or more long-standing. At follow-up, the patient below had been suffering from persistent headaches for the

previous fortnight and had consequently nominated health as a cue. At baseline he had not referred to health:

II#32: Yes, it's kind of wearing me a bit thin at the moment so that's why it's come up.

(Male / 46 years old / referred / IT project manager)

In contrast, the next patient described how her experience of financial concerns several years previously, continued to affect her evaluation of finances:

II#38: Again you know if everything is going fine you don't think about finances and most of the time it has been but we've had some hiccups in that my husband has been made redundant a few times from different jobs.

(Female / 47 years old / referred / teacher)

For this patient then, an earlier increase in weighting for finance had possibly resulted in a stable rating of cue importance, and indeed 'finance' was nominated as a cue at both assessments. Another patient described how the knee problem from which he was now largely recovered, had continued to exert an influence on his cue weighting:

II#19: ... you only kind of realise the importance of something when there's been a problem with it and in a sense of health, for me to realise how, not debilitating it is, 'cos it hasn't been for me, but you realise how even the small niggly things can like effect so many different things – not so much compound the problems but they have such a wide effect that I mean if someone had said to me "Oh, you're going to have a problem with your knee and you're going to be in discomfort like running or doing sport, its going to like cause an upheaval in so many different areas". You think, "Oh, no way, I just won't be able to play sport"

(Male / 24 years old / listed / medical student)

His newly developed appreciation of health had persisted beyond his immediate experience of health problems. Furthermore, an added impetus appeared to have been his surprise at the impact of the knee problem across the breadth of his life. It is possible that because his *expectations* about the

consequences of his knee problem had been more limited than the reality, the impact on his reprioritisation was subsequently greater.

For other patients changes in weighting may have been more transient and even, for them, predictable. Thus, for patient #4, current financial restrictions reduced her opportunity for socialising, and hence she had downgraded its importance. That it would resume eventually and become important again was reflected in her comments. Therefore, this reprioritisation was likely to be more transient:

I#4: Socialising, I suppose once everybody comes into a bit of money now and we all start going out again, it will be important.
(Female / 24 years old / listed / student)

Health changes: The impact of changes in health was of particular interest, and most often this meant deterioration in health. Experience of health problems often increased the importance of this cue to individuals. The following three patients described this effect on their perceptions, including how the importance of health had simply been taken for granted during times of good health. Patient #38 also reflected on how the importance of health may accelerate with the ageing process. For patient #20, health had receded in importance now he was experiencing fewer problems:

I#38: ...I mean if you're healthy you don't think about it, and if anything crops up, then you start thinking about health, yeah.

MR: Right, okay.

I#38: And probably getting older makes it part of your life as well.
(Female / 47 years old / referred / teacher)

I#39: So, I think that health is the most important thing you've got and until you don't have a good run of health – you know – until your health goes down you don't realise how important it was before.

I#39: It depends what's going to go wrong with our health as to how important it is really doesn't it?
(Female / 50 years old / referred / local government administrator)

II#20: ... I was more concerned about the health than I am now because I'd just recently broken the other leg which is probably a factor into why I haven't mentioned it this time, because its not really bothering me at the moment.

(Male / 29 years old / listed / security guard)

Comprehension, correlation and direction of relationship: As Moons and colleagues noted there are occasionally comprehension difficulties with SEIQoL completion.⁵⁹ In the current study, when answering questions about cue level, some patients responded by reference to the importance of the cue. It is possible that this was simply to convey a fuller picture of the cue. A second explanation may be that the patient was drawing attention to their own perception of a relationship between cue level and weight. Alternatively cue level may simply have been confused with the weight and, therefore, their completion of SEIQoL may have been partly erroneous. When asked about her rating of the *friends* cue level, the following patient described the value (weight) she assigned to other cues:

MR: ... at the moment you're saying that it's not the best it possibly could be?

II#18: Not really, no, but that's my own choosing.

MR: Right, okay, okay.

II#18: But my health and my family is most important.

(Female / 64 years old / referred / retired)

Whilst it is apparent that the level of a quality of life domain may have influenced its provision as a cue, there was also some evidence that weight accorded to a cue may also have affected the reported level (i.e. judgement and reporting of cue level is dependent upon the cue's perceived importance). An example of this was provided by the following patient, who spontaneously described the rationale for his baseline rating of *work* level (which was 56/100):

II#20: Work is in-between 'cos like I say it's not important

(Male / 29 years old / listed / security guard)

The actual relationship may have been more dynamic, for this patient at least, as he subsequently described how the importance of *work* resulted from his inability to work:

I#20: Work at the moment is down at the bottom of the scale [weighting] because I can't a – do any work
(Male / 29 years old / listed / security guard)

He ranked the importance of this cue lowest of all the five cues and this may help to explain why he rated this cue just above average, when at the time he was actually unable to work. When asked to describe why he rated the cue thus, he described how he was unconcerned about the prospects of returning to work and his confidence about finding an alternative employer if necessary:

I#20 Work – it's not really a priority ... all my friends, we all work in the same line of business, same work, you know they said come and work for them, come and work for them now, their bosses will take us on without a problem.
(Male / 29 years old / listed / security guard)

Cue weight: a choice or imposition? The weighting expressed by respondents may have been determined by several factors, including the level (or change in level) of the cue itself. What though does the provided weight represent – is it an *observation* by the patient of the relative contribution of that cue to their overall life quality, or a statement of how important they feel each cue *should be* in their life. Stated differently, to what extent do respondents consider the importance of a cue their active choice or simply something imposed by their life circumstances? Some data arose from the interviews. For example, the following patient's low cue weighting for the son with whom she was in dispute, appeared to be a deliberate choice upon her part:

I#29 He's got to be last 'cos he's showing no compassion at all.
(Female / 54 years old / listed / catering worker)

Whilst she may have only limited influence over her son (and, therefore, the level of that particular cue), she was able to exert some control over the importance she attached to it. Thus, this may represent an adaptive process which served to minimise the effect of the problem upon her quality of life. However, for the next patient who was considering the importance of *health*, there appeared to be little room for manoeuvre. Health was important to her because it was pre-requisite to fulfilling her family care obligations:

MR: Okay and then your health. You've already mentioned why that's important to you.

II#35: That's right, because in order for me to continue keeping that going.

MR: Mm

II#35: That has to be a good factor in my life as well.

(Female / 63 years old / referred / retired social worker)

10.4 Discussion

Overview: A secondary analysis of quantitative and qualitative interview data explored the relationship between the level of SEIQoL cue and its weight (i.e. between how satisfied a patient was with a cue, and how important he/she considered the cue to be). Univariate analysis found evidence of a small positive correlation between cue level and weight. There were, however, some instances from the quantitative (and qualitative) analysis where the relationship appeared to go in the opposite direction. The strength of the correlation was not so great as to suggest that patients confused level and weight when forming their response. Descriptive qualitative analysis provided additional support for a relationship between cue level and weight. There was also evidence which indicated how change itself in cue level may precipitate cue weight change, and highlighted the place of health as a quality of life cue. Finally, further exploration of the meaning of cue weight identified instances when cue weight may itself influence perceived cue level, and occasions when the weighting of a cue represented an active patient choice.

Moons and colleagues reported a significant correlation between SEIQoL cue level and weight ($r=0.26$, $p<0.001$) which is similar to that for cues elicited at

follow-up, and cues from listed patients in this study.⁵⁹ They took this to be evidence of a low to moderate correlation, thus supporting the validity of the internal structure of the assessment. In the current study, there was no significant correlation between cue level and weight for baseline cues, nor for cues from waiting list patients. Therefore, there was no suggestion that patients confused weight and level. A change in the relationship between cue level and weight from baseline to follow-up, may have resulted from response shift if, for example, patient priorities had become more influenced by their health status. Similarly, patients with more long-standing health problems (i.e. listed for arthroscopy) may also be more influenced by their health status, than those with a more recent onset.

Limitations of the analysis: Conclusions about differences between cues derived at baseline and follow-up, or between arthroscopy and waiting list patients should be tempered by recognition of limitations in the current analysis. Whilst calculating correlation coefficients for cue level and weight for each patient accounted for within-person clustering, the inter-relationship between potential confounding variables should also be addressed. The results remain suggestive rather than conclusive, but do offer a direction for future analyses. Campbell and Whyte, in their SEIQoL study of patients with cancer, noted the good correspondence between cue weight and level for one profiled patient but do not suggest why they should be correlated.²³⁶ In contrast, Moons and colleagues proposed their criterion for the internal assessment structure as a validation check against patients confusing level and weight.⁵⁹ In this study, this criterion was satisfied regardless of study group or timing of interview, as in no instance was the correlation very large.

The primary correlational analysis assessed 345 cues and was adequate to demonstrate a significant statistical relationship between cue level and weight. It also indicated differences between individuals in the extent and direction of relationship between these two variables. However, a larger sample of respondents may allow further analysis of how this relationship varies with potential correlates such as disease trajectory and socio-demographic

variables whilst for controlling for multiple observations from each individual respondent.

The relationship between cue level and weight, and response shift: The descriptive analysis provided examples of patients linking cue level and weight, usually describing a casual relationship. This relationship could be both positive and negative with, for example, high levels associated with both high and low weights. If the weight attached to a cue was partly determined by its ambient level then reducing weight when a cue was problematic, and increasing weight when it fared well, may be examples of beneficial adaptation. Of course, the extent to which this is possible may vary with the nature of each cue. Certain cues may be harder for individuals to downplay when things are not going well, for example, health. Indeed, adversity serving to increase the importance of cues seems natural, and is what Aristotle was referring to in the quote of the start of the chapter.²⁷⁴

Replacement changes in the SEIQoL cue profile (see chapter 8) are the clearest form of re-conceptualisation response shift. Of all the potential areas of life influencing an individual's quality of life, they will perceive some to be of greater importance and these will be selected within the SEIQoL assessment. If the level of a candidate domain influences its weighting, this will increase its chance of nomination. Thus, as the level of different areas of life fluctuates over time, an individual may tend to focus upon, and select, those which attain a minimal level of satisfaction. Therefore, whilst the cue profile changes, and patients re-conceptualise, their observed overall quality of life level may remain relatively constant. Patients may achieve quality of life homeostasis by weighting and selecting cues partly on the basis of their ambient level.

Cue distributions: The distribution of cue level and weight scores is worthy of comment. For the former, there may be evidence of end-aversion bias and incomplete use of the whole scale by the selection of certain (modal) points along the scale. Patients may, therefore, discriminate less finely in their response than the scale actually allows. The distribution of weight scores is different and reflects the nature of their elicitation. Unlike cue levels, cue

weights are constrained to sum to unity - the aggregate weights must total 100.¹³⁵ As O'Boyle commented, allowing free weights may offer advantages for exploring re-prioritisation response shift.

To weight items or not?: Individualised approaches to assessment such as SEIQoL-DW specifically allow weighting of quality of life items to reflect individual preferences. Standardised nomothetic approaches may also weight items although these would be either theoretically or empirically derived and applied uniformly across all respondents. Streiner and Norman reviewed the value of weighting items.⁶ They concluded that weighting adds little when there are a large number of items (40+) and when the items are relatively homogenous. Thus, for the latter case, very similar items are likely to attract similar weights from respondents (i.e. the weights will be within a narrow range). In contrast, a SEIQoL index score is derived from only five individually nominated items (cues), which themselves are likely to be heterogeneous.

The theoretical impact of re-prioritisation upon derived index score was demonstrated in chapter 8. That individuals may assign different weights (importance) to different life domains is central to the concept of response shift re-prioritisation. Furthermore, the impact of weighting or not SEIQoL cues is shown in a theoretical example in table 10.4.1. Where levels are equivalent for each cue (I), weighting cues adds nothing to the overall index score – both total 75. Where cue levels vary, different index scores result (II). In this example, the difference between weighted score (58.5) and un-weighted score (42) is 16.5.

Table 10.4.1 Effect of weighting cues on SEIQoL index score with (I) equivalent cue levels and (II) different cue levels

Cue	I: Equivalent cue levels		II: Different cue levels	
	Level	Weight	Level	Weight
A	75	.40	80	.40
B	75	.30	60	.30
C	75	.15	40	.15
D	75	.10	20	.10
E	75	.05	10	.05
Index score:				
Un-weighted	75		42	
Weighted		75		58.5

10.5 Summary

Although authors such as Moons and colleagues have explored the relationship between SEIQoL cue level and weight, their interest has been mainly in what this reveals about the validity of the assessment itself. That patients adequately differentiate between these two concepts is indeed important. However, the current study has explored further the relationship and provided a possible mechanism by which response shift may result. Specifically, changes in cue level may, in part, determine the priority accorded to that cue by the respondent. Whilst the focus has been upon cue weight, and hence re-prioritisation, it is also clear that re-conceptualisation may also result from the same mechanism. Focusing attention away from a problematic area of life may be adaptive in dealing with the problem and result in a response shift. Focusing attention towards the problem may also result in response shift but may result in a lower evaluation of quality of life. The direction of the impact appears variable, for example, current health problems may serve to increase or decrease the relative importance of health. Possible factors influencing the relationship between cue level and weight, for example, study group and timing of interview, may be useful areas for study in the future. A key general aim of such work would be a fuller understanding of the meaning and dynamism of patient priorities as expressed in SEIQoL cue weighting.

Preface to chapter 11

The current chapter concludes the empirical response shift analysis presented in the thesis. The analysis has largely been presented at an individual level. However, group level approaches for analysing individualised data (such as generated by SEIQoL) have been used in other studies, although not necessarily to evaluate response shift. In designing the studies of this thesis, such approaches were considered but particular challenges identified. Therefore, the following chapter presents a structured review of approaches to quantitative analysis of SEIQoL. The review describes studies that have taken this approach and uses data from the current response shift interviews to illustrate some of these challenges.

Chapter 11: Group level assessment of re-conceptualisation and re-prioritisation: a commentary and some recommendations

Abstract: *Whilst SEIQoL is an individualised measure, it has been used for group level analysis in both cross-sectional and longitudinal designs. One potential application of SEIQoL in longitudinal designs is to evaluate response shift. This chapter reviews the quantitative use of SEIQoL-DW data and examines some of the particular challenges of these approaches in the presence of response shift. These challenges primarily relate to basic classification and coding of cue data and the validity of subsequent comparisons. Such challenges are presented using examples from the current study data set. The need to systematically classify cue data prior to quantitative analysis is emphasised. Recommendations for future practice are provided within the discussion section.*

11.1 Introduction

Individualised quality of life measures have been used for quantitative group level analysis, and using SEIQoL to assess response shift was initially considered as a secondary analytic option within this thesis. However, several methodological issues related to such group level analysis become apparent in the planning phase, and thus the study was ultimately designed to focus upon individual level change. Nevertheless, this chapter reviews some of the issues identified, with particular reference to empirical studies which have employed using SEIQoL for making quantitative comparisons. Furthermore, data from the thesis interviews, and insights from the typology of change presented in chapter eight, are used to demonstrate the challenges of using individualised outcome data to assess response shift. At the end of this commentary, some recommendations for using SEIQoL data for making quantitative group comparisons are provided.

11.2 Using aggregate SEIQoL data in quantitative analyses

The methods review by Schwartz and Sprangers, and the work of O'Boyle and colleagues identified examples where aggregate data from an individualised measure (e.g. SEIQoL) had been both proposed and utilised within quantitative analyses to assess response shift.^{11 135} More broadly, further examples where simple group comparisons have been made based on individualised outcome measures were apparent (for example, using SEIQoL index scores). Indeed, the SEIQoL administration manual describes how, for the purpose of grouping data, index scores may be presented.¹⁹²

For the purposes of this commentary, potential quantitative approaches to analysing SEIQoL data were identified (i.e. not just for the assessment of response shift). Data may be summarised at the level of index score (the sum of the product of individual cue levels and weights), as well as at the level of individual cues. For the latter, data may relate to frequency of cue elicitation; reported levels; and cue weights. Analysis may be merely descriptive in nature (for example, mean index scores or frequency counts of nominated cues) or inferential. Furthermore, analysis may result from either a cross-sectional or longitudinal study design, with the latter being most applicable to the study of response shift. These various approaches have been summarised in table 11.2.1. The reference list for the SEIQoL administration manual, and the Medline database were purposively reviewed to identify studies providing examples of the various analyses outlined in table 11.2.1. The scope of the search was therefore, illustrative rather than comprehensive.

Table 11.2.1 Quantitative approaches to analysing SEIQoL data

Assessment level	Study design	Statistical approach	Comparison	Statistic	Examples
Index score	Cross-sectional	Descriptive	-	Mean	Campbell ²³⁶
	Longitudinal	Inferential	Unmatched	t-test	Bromberg ²⁷⁵
		Descriptive	-	Mean Median	Bayle ²⁵⁶ Smith ⁵⁸
		Inferential	Matched	Paired t-test Wilcoxon SR	O'Boyle ²⁰¹ Smith ⁵⁸
Cue frequency	Cross-sectional	Descriptive	-	Counts / %	Hickey ¹⁹¹ Mountain ²⁷⁶ Frick ²⁷⁷
		Inferential	Unmatched	χ^2 test	Wettergren ²⁰³ Clarke ²⁷⁸ McGee ¹⁸⁸
			Matched ⁽ⁱ⁾	McNemar	O'Boyle ²⁰¹
	Longitudinal	Descriptive	-	χ^2 %	Bayle ²⁵⁶
		Inferential	Matched	χ^2 test ⁽ⁱⁱ⁾	Browne ¹⁹⁰
level	Cross-sectional	Descriptive	-	Mean / SD	Wettergren ²⁰³
	Longitudinal	Inferential	Unmatched	t-test	Wettergren ²⁰³
		Descriptive	-	Mean	Bayle ²⁵⁶
		Inferential	Matched	Paired t-test ⁽ⁱⁱⁱ⁾	O'Boyle ²⁰¹
weight	Cross-sectional	Descriptive	-	Median	Campbell ²³⁶
	Longitudinal	Inferential	Unmatched	^(iv)	Waldron ¹⁹³
		Descriptive	-	Mean Mean difference	Bayle ²⁵⁶ Browne ¹⁸⁹
		Inferential	Matched	Weighted κ statistic ^(v)	Browne ¹⁸⁹

Notes for table 11.2.1:

- (i) Although O'Boyle's study was longitudinal, comparison used baseline cues*
- (ii) Browne describes ' χ^2 comparisons' for various cross-sectional (unmatched) and longitudinal (potentially matched) comparisons but does not specify use of McNemar χ^2*
- (iii) Comparison was of cues identified by order of weighting, regardless of cue label*
- (iv) Actual test used by Waldron not specified. Comparison was of weights for two cues provided by the same patients*
- (v) Browne converted weights to ranks to assess stability (agreement) of weighting procedure (7-10 day test-re-test interval)*

Index level analysis: Descriptive analysis of SEIQoL index data has involved presentation of mean and median values, for both cross-sectional and longitudinal designs.^{58 236 256} Inferential analysis of index scores has included the t-test (for unmatched cross-sectional data), and the paired t-test and Wilcoxon signed rank test for matched longitudinal data.^{58 201 275} Thus, Bromberg and colleagues compared patients with amyotrophic lateral sclerosis and their carers.²⁷⁵ The higher index scores for the patient group was found to be significantly different from that of the carers using a two-tailed t-test. Similarly, using a paired t-test in a longitudinal design, O'Boyle found significant improvement in quality of life scores in patients following hip replacement surgery.²⁰¹

Cue level analysis: At the analytic level of the cue, three components are considered here: frequency of cue nomination; satisfaction level; and weight of cue. For the first component, frequency counts and proportions naturally form the basic level of descriptive analysis for both cross-sectional and longitudinal data (table 11.2.1). There are many examples of this level of presentation, including Mountain and colleagues who described how frequently elderly medical patients nominated each of thirteen SEIQoL-DW areas.²⁷⁶ Similarly, Hickey and colleagues presented the frequency with which SEIQoL-DW life areas were nominated by HIV positive patients, and the proportion of the sample represented by each cue area.¹⁹¹ Whilst their study included an age-matched control group, there was no formal statistical comparison of cue content. Finally, in a longitudinal study, Bayle and

colleagues presented the proportion of patients nominating specific cues before and after hip arthroplasty but again this was not tested statistically.²⁵⁶

However, there are several instances where statistical association was formally tested, using either the χ^2 test or the McNemar χ^2 test. Wettergren and colleagues, compared the proportion of two un-matched study groups (patient and healthy control) nominating individual cues (e.g. *family, personal health*) using the χ^2 test.²⁰³ Clarke and colleagues categorised patients with amyotrophic lateral sclerosis according to functional impairment (above / below median score) and compared frequency of cue nomination.²⁷⁸ They used the χ^2 test in a cross-sectional unmatched group comparison. It is also worth noting that even when formal comparison is not undertaken, indirect comparisons are often made, for example, within discussion sections.²⁰⁴

Where data are matched, the McNemar χ^2 test has been used. O'Boyle and colleagues compared patients undergoing hip replacement and healthy controls (matched by gender, age and socio-demographic status) in frequency of nominated cues.²⁰¹ Although the study was longitudinal, this comparison only used baseline cues and change in frequency of cue nomination over time was not assessed statistically.

Browne and colleagues reported the proportion of healthy elderly respondents nominating the same cues at two time-points over a twelve-month period.¹⁹⁰ Although the longitudinal data were matched, differences across assessments were formally tested using a χ^2 test. However, the McNemar test is more suitable than the χ^2 test for intra-individual comparisons (i.e. pre-post observations of the same individual) as it accounts for the same sample being used on both occasions.

Descriptive analysis of satisfaction level for individual cues has been reported using cue means.^{203 256} Formal comparisons have used t-tests and paired t-tests for cross-sectional (unmatched) and longitudinal data respectively.^{201 203} Wettergren and colleagues compared long-term survivors of Hodgkin's

lymphoma with a control group in terms of their satisfaction with different life areas, but found no significant differences.²⁰³ O'Boyle and colleagues compared levels of satisfaction pre- and post-operatively in a sample of patients undergoing hip arthroplasty.²⁰¹ Interestingly, their comparison was of cues identified by their order of importance, rather than cue label.

Descriptive analysis of cue weight has included mean and median presentation, as well as presentation of mean difference in cue weight across assessments.^{189 236 256} The latter analysis by Browne and colleagues was part of a method comparison study assessing the stability of weights derived either by judgement analysis or direct estimation (i.e. SEIQoL versus SEIQoL-DW). A comparison of weights assigned to two cues (*family* and *health*) amongst a sub-group of patients with advanced cancer was presented by Waldron and colleagues, although the test is not specified.¹⁹³ In Browne and colleagues' longitudinal method comparison study, weights had been converted to ranks for analysis.¹⁸⁹ Agreement between baseline and follow-up weights (from 7-10 days later) was assessed using the weighted κ statistic.

Mean number of cue changes: In addition to the cue level analysis described in table 11.2.1, descriptive analysis has addressed the mean number of cue changes over time. For example, Browne and colleagues reported an average of 1.1 domain changes described by healthy elderly respondents over the course of a year.¹⁹⁰ Similarly, Bayle and colleagues reported five patients providing the same cues before and after total hip arthroplasty, and the remaining 25 patients changing on average 1.6 items.²⁵⁶

Correlation and other analyses: SEIQoL data may also be used in various correlation analyses, particularly at the index level where the relationship between SEIQoL score and other outcome measures may be of primary interest. Thus, Prince and Gerber described the relationship between SEIQoL score and two other measures of life quality, in patients with serious mental health problems.²⁷⁹ Classification based upon SEIQoL completion may also form the basis for analysis. In one sample of patients with diabetes, Wagner

and colleagues compared the ages (and other socio-demographic variables) for two sub-groups formed according to their nomination of one or other aspect of a single diabetes-related cue.²⁶¹ In this example, the validity of the process for classifying patients into one or other group was obviously critical.

Assessing change using the Kappa statistic

Assessing change over time with paired categorical data using the χ^2 test would be inappropriate as this tests association.²⁸⁰ An alternative approach (used by Browne and colleagues) would be the use of the κ statistic, which assesses level of agreement above that expected by chance.^{189 281} It can be used where one is assessing the presence or absence of a cue at two time-points (i.e. a 2x2 table). Guidelines for interpreting kappa values are available (e.g. moderate agreement being 0.41-0.60, very good agreement being 0.81-1.00). One consideration though with the use of kappa is that its value is affected not only by agreement but also the prevalence of the attribute being measured. Therefore, caution should be taken when comparing cues of varying prevalence.

11.3 Administration and cue elicitation

The various ways in which SEIQoL has been administered has implications for its quantitative analysis. Identifying domain re-conceptualisation in a longitudinal study requires new cues to be elicited at follow-up time-points. The SEIQoL-DW administration manual allows for this, and also suggests that baseline cues, if different, are provided for the respondent at follow-up and the process repeated to enable direct comparison across time-points.¹⁹² O'Boyle and colleagues described this procedure in a study of patients undergoing hip replacement, in which the discrepancy between baseline and follow-up cues was noted.²⁰¹

SEIQoL requires respondents to nominate five cues, a figure originally set due to the difficulties of combining larger numbers of cues when forming a judgement.¹³⁵ Occasionally, in practice, fewer cues may be provided.

Conversely, Wettergren and colleagues described the use of an amended procedure which permitted an unlimited number of cues.²⁰³ In such circumstances, comparing frequency counts of respondents nominating individual cues, either in cross-section or longitudinally, would still be appropriate. However, comparing data derived from conventional administration (which may restrict the provision of less important cues) with data produced using unlimited cues would seem inappropriate.

Further modifications to the administration of SEIQoL-DW have been reported with the aim of simplifying and facilitating response. Wagner and colleagues simplified the language used in their study of children and adolescents with diabetes, and also asked respondents to feedback the provided instructions (i.e. to 'retell') in order to verify comprehension.²⁶¹ The response format for reporting cue levels has also been modified. For example, Mountain and colleagues provided seven categories ranging from 'best possible' to 'worst possible', with each allocated a value for computation.²⁸² Finally, in addition to the original and direct weighting methods for SEIQoL, the use of ranks for quantifying domain importance has been used by Tovbin and colleagues.²⁰⁴ Comparisons across studies may be confounded when differing response formats for either cue level or weight have been used.

Classifying and coding cues

The validity of quantitative comparison of SEIQoL cue data is also affected by how cues are initially classified and data coded. The SEIQoL administration manual refers to the variety of content that may be assigned a common cue label and, therefore, the importance of summarising cue meaning during administration.^{135 192} This is important for two reasons. Firstly, it identifies whether the same cues are being addressed at subsequent re-assessment. Secondly, it enables summation of cues from several respondents for grouped presentation. However, the manual offers no further guidance on how cue data should be classified and coded for group level comparison.¹⁹²

The idiosyncratic nature of meanings included under cue labels means that a single label within a study may incorporate a variety of themes.¹⁹⁰ This would

suggest a move beyond such labels in exploring the nature of change within an individual, and emphasises the role of individual level analysis. At the very least, the individual nature of SEIQoL cue nomination and labelling creates a challenge for group comparisons operating at the cue level. Analysis of change of cue content (and by implication cue level and weight) requires an appropriate method for classifying cues and aggregating data.

Developing a classification framework to 'make sense' of the data is a necessary step and requires the identification of recurring regularities – i.e. what fits together - in forming categories (termed convergence).¹⁸⁵ Resultant categories may be judged according to internal homogeneity (that data units within a single category hold together) and external homogeneity (differences between categories are distinct). Subsequent divergent strategies, including saturation of categories and deviant case analysis, help to verify the developing framework. Specific formulations for analysing such data, for example, methods of constant comparison; open, axial and selective coding have also been described.^{210 249}

Issues in classifying SEIQoL data: A number of key issues in classifying SEIQoL data are apparent from the literature (table 11.3.1). Thus, Lindblad and colleagues have questioned who should be responsible for categorisation (e.g. the patient or the researcher).¹⁹⁷ Adopting the respondents perspective reflects the anthropological 'emic' analytic approach and the consequent use of indigenous categories and terms has been labelled 'in vivo' coding.¹⁸⁵ This contrasts with the 'etic' approach which uses researcher-assigned labels, and allows the use of the analysts own insights.²⁵⁰ Furthermore, developing a classification framework may be purely inductive, or instead refer to previous classifications. For example, Bayle and colleagues grouped cues according to 'classic' categories provided in the SEIQoL-DW administration manual (table 11.3.1).²⁵⁶ Other a priori frameworks exist, including those developed by Bowling and by Wettergren and colleagues.^{203 283}

Table 11.3.1 Issues in classifying SEIQoL data

<i>Issue</i>	<i>Examples</i>	<i>Comments</i>
Basis for framework classification	Bayle ²⁵⁶	Should classification be respondent or researcher led? Classification with reference a priori framework
Use of 'miscellaneous' or 'other' category	Gribbin ²⁸⁴	Rationale for inclusion in 'other' category not made explicit. Some infrequently nominated cues remained as separate categories
	Ramström ²⁸²	In addition to miscellaneous category other categories with only two patients nominating the cue (e.g. Independence).
Process of classification made explicit	Wagner ²⁶¹	Rationale for including response within categories described (e.g. for 'family' if term 'family' or 'parents' used)
	Moons ⁵⁹	Transcription and sorting of responses described
Specification of sub-headings	Clarke ²⁷⁸	Sub-headings not used in quantitative analysis
	Wagner ²⁶¹	Sub-categories formed basis for group comparison

A feature of some studies is the use of an 'other' or 'miscellaneous' category.^{188 282 284} In a study of patients with cystic fibrosis, Ramström and colleagues included a 'miscellaneous' category which incorporated single responses related to 'travelling', 'God' and 'food'.²⁸² Idiosyncratic SEIQoL cue labels may indicate the use of such categories, but makes comparisons difficult. Furthermore, a thematic analysis of cue content moving beyond the respondent's own cue labelling may identify broader, but cohesive, categories which could satisfactorily accommodate such data.

Classifying data in practice: In summarising SEIQoL cue data, the process of developing a classification framework has generally neither been attempted, nor well described. Tovbin and colleagues, for example, presented the five most frequently nominated life domains in a study of haemodialysis patients (including *economic* and *leisure*), with no description of how these categories were formed.²⁰⁴ In a cohort study of people with HIV/AIDS, Hickey and colleagues presented a detailed listing (with occasionally more descriptive labels), but again, with no formal category definition.¹⁹¹ Other

authors have explicitly grouped elicited SEIQoL domains into 'aggregated cues' but still not described how this was achieved or provided category definitions.^{195 243 277}

However, a small number of studies did follow explicit procedures for classifying SEIQoL cues. Moons and colleagues sorted transcribed cue descriptions into clusters which were then subjectively labelled.⁵⁹ Wagner and colleagues described the rationale for allocating responses to categories and provided examples of the process.²⁶¹ For example, responses were coded as *family* if the term 'family' or 'parents' was used by respondents. Their data handling also demonstrates a level of sophistication whereby domains were further sub-divided according to content. Thus, under a *diabetes* domain, two sub-categories were distinguished according to whether content referred to *self-care behaviours* or *living well with diabetes*. Similarly, Clarke and colleagues delineated three sub-categories of a single disease-related category in a study of patients with amyotrophic lateral sclerosis (ALS).²⁷⁸ In a second study of ALS, Neudert and colleagues classified as *existential* all cues which related to personal growth, transcendence or purpose / meaning of life.²⁸⁵ Even if such sub-categories are not addressed in further analysis, their provision enhances clarity of category content and enables transparent coding.

Coding data: The existence of a classification framework still requires data to be reliably coded to it, and again this process is usually not described. Thus, a classification of life areas for patients with leukaemia and lymphoma developed by Montgomery and colleagues, included categories such as 'awareness/positivity' but no description of how responses were coded to such a heading.²⁸⁶ Despite the classifying procedure described by Moons and colleagues, they present no information on coding verification or rater agreement.⁵⁹ Furthermore, it is unclear how many categories were used and the extent to which they were mutually exclusive, and coped with the potential coding dilemmas.

Coding data to a classification framework usually requires a structure that is exhaustive with mutually exclusive categories. This is especially important if one is to count and compare occurrences of specific cues. Frick and colleagues noted the valid occurrence of multiple cues (per patient/assessment) assignable to the same aggregate cue header.²⁷⁷ They treated this as three occurrences of that cue in their descriptive analysis. It is possible that a more refined classification with a larger number of categories or sub-categories may reduce such multiple occurrences.

11.4 Using aggregate SEIQoL data to assess response shift

Much of the discussion above has focused upon cue-level, rather than index score level, analysis. In response shift assessment, this is mostly relevant to re-prioritisation and re-conceptualisation. Whilst several studies described above assessed change over time, only a few explicitly addressed response shift. For example, the study by Browne and colleagues of healthy adults was presented by O'Boyle and colleagues as an example of how SEIQoL may be used to assess response shift, but they noted that such an assessment was not an aim of the original study.^{135 190}

O'Boyle and colleagues described the role of relative and free weights for assessing re-prioritisation response shift.¹³⁵ Whilst SEIQoL weights are constrained to sum to unity, they considered that independent weights may be more appropriate for assessing change in values. This may be an argument for focusing more on the patient description of change in an individual level analysis, rather than relying upon either absolute change in cue weight or change in relative cue ranking.

The nature of response shift changes identified in chapter eight has some implications for how cue data should be managed and coded in group level analysis. Four scenarios of particular relevance are presented in table 11.4.1. In the first (p#22), a single cue at baseline incorporated content that could be coded under one of two domain categories (*health* or *work*). In this instance,

the patient had combined or aligned two constructs which otherwise could have been assigned to separate categories. Indeed, at follow-up, *work* and *health* both re-appear but this time as distinct cues. This type of change was described earlier as a re-formulation (separation) response shift. In a second example (p#23), *being healthy* is a baseline cue and at follow-up both *health* and *keeping fit* are provided. In this second scenario, the single baseline cue was perhaps a more logically cohesive cue compared to the first scenario, but has nevertheless still undergone a reformulation and separation of its constituent parts.

Table 11.4.1 Problematic coding scenarios

Coding scenarios	Baseline cues	Follow-up cues
P#22	<i>combined health and work cue</i> money family sport social life	<i>work</i> <i>health</i> sport family social life
P#23	Family activities Marriage <i>Being healthy</i> Children's education Working (shifts)	<i>Health</i> Family Money <i>Keeping fit</i> Children's education
P#39	<i>health (i.e. physical)</i> work family gardening home	<i>health (inc. mental health)</i> family finance work home
P#32	Family Financial security <i>Job security</i> <i>Job satisfaction</i> Wales rugby	Family <i>Work</i> Finance Health Wales rugby

This creates a number of difficulties for quantitative comparison. For p#22, should the baseline cue be counted as an example of a *health* or *work* cue (or even a third category of cue) when comparing counts of baseline and follow-up cues? Choosing to code the one original cue as two separate cues would initially appear to be a reasonable strategy. It would not, for example, affect

assessment by kappa, which is simply concerned about presence of absence of a cue. However, including the original 'combined' cue separately under two specific categories may misrepresent the intention of the respondent. It may also not be possible if the description of the cue is too general to allow such dis-aggregation (as may be the case for *being healthy*, p#23). Furthermore, it inadvertently allows six cues at baseline, whilst retaining five at follow-up. This confounds assessment of number of cue changes per patient. For example, for p#22 it is possible that *money* (or another cue) may have been provided at follow-up if a sixth cue had been requested. Thus, apparent change may actually be an artefact of data collection and management.

The patient in scenario three (p#39) provided a cue coded as *health* but which was narrowly focused upon physical functioning. Subsequently, they described health with a broader range of components which explicitly included mental health. This is an example of what was referred to earlier as a re-formulation (expansion) response shift change – the scope of the original cue has been extended.

How is this problematic? After all, a single cue has been provided on each occasion, both of which could simply fall under a general *health* heading. However, there has been a change of emphasis on the part of the respondent and it would seem appropriate to recognise this in cue categorisation. The validity of assuming that baseline and follow-up cues are equivalent for the purpose of group comparison could at least be questioned. Thus, depending upon the nature of the cue description provided by patient #39, the follow-up cue could be divided into two cues representing physical and mental health. Analysis would show stability in physical health at a group level comparison, but change in mental health.

What should happen, though, if the original cue described simply *general health*, and at follow-up, separate mental and physical health were provided? With which follow-up health cue should the baseline health cue be compared? This dilemma is also demonstrated in the fourth scenario (p#32). Two separate baseline cues (job security and job satisfaction) could both be coded

under a single work category, similar to the multiple cue situation described by Frick and colleagues.²⁷⁷ In this instance, as different facets of work are being emphasised by each cue, perhaps a more refined categorisation could accommodate these cues. However, it is highly likely that essentially equivalent cues could be provided, for example the description of two hobbies which are truly multiple occurrences of the same cue. How should baseline and follow-up weights for work be compared for p#32, if at all? Given such multiplicity should the weights for job security and job satisfaction be aggregated or averaged? How meaningful is it to compare the weights for a health cue which emphasises physical health with a more general health cue which explicitly includes mental health? The validity of group level comparison rests upon the adequate answering of such questions.

In summary, across each of these scenarios the way cues are framed by respondents, or categorised in analysis may covertly or overtly change the actual number of cues being compared from baseline to follow-up. A finer categorisation of cues may facilitate valid comparison of cues but may produce a classification framework that is somewhat unwieldy and may still fall short. This is perhaps likely where categories and sub-categories may be used. A less refined classification, with fewer categories may present less coding and comparison difficulties but is likely to be too insensitive to the more subtle response shift changes that may occur. More substantial response shift change at the level of cue profile where one unambiguously distinct cue is clearly replaced by another may be suitable for group comparison. However, in an assessment context where individuals are free to define idiosyncratic cues which are more personally valid, such ambiguity is likely - indicating a more individual approach to analysis.

11.5 Recommendations for quantitative group comparisons

This review makes the fundamental point that cues need to be freshly elicited at each administration to allow assessment of cue content change. Whilst retaining baseline cues still permits assessment of change in weights,

requiring the patient to rate (level and weight) the original cues may generally invalidate the assessment. A pre-requisite for quantitative group comparison (whether for the determination of response shift or more generally) is a systematic categorisation of respondents' cue data. This should incorporate the individual's unique perspective, but also draw upon insights of the analyst. Reference to an existing classification framework may be useful when categorising cues, but the framework should incorporate all novel cues elicited in the study.

To facilitate transparency and repeatability, the final classification should be supplemented by at least a simple coding frame with accompanying definitions and coding instructions. This may involve multiple raters for at least a proportion of coded cues. A process of resolving disagreement should be specified. It should be remembered that valid assessment of the number of changed cues across assessments may be greatly and covertly affected by coding decisions which may imbalance the number of domains being compared. Therefore, such assessment should be treated cautiously.

Once data have been adequately classified, quantitative analysis may proceed. Nomination of particular cue domains to be tested before analysis would reduce multiple significance testing and the possibility of Type I error. Differences between the proportion of respondents nominating each domain at baseline and follow-up should be assessed by treating the outcome as binary (presence or absence of cue). As this involves a pre-post comparison of matched cases with a binary variable, agreement may be assessed using the kappa statistic. Patients not providing cues at either assessment (due to difficulties with the assessment or failure to follow-up) should be excluded from the analysis.

Group differences in assigned weights for individual cues will be assessed using the paired t-test (or if indicated by the distribution of data, the Wilcoxon signed rank test). Valid assessment at the cue level (frequency of cue nomination, cue level and cue weight) is dependent upon the equivalence of the cues being compared, which at least requires careful scrutiny.

11.6 Summary

Quantitative group level analysis of SEIQoL data was considered in the design stage of the response shift study, but a number of issues related to such analysis were identified. In this chapter I have presented a structured review of how such data have been previously analysed using descriptive and inferential statistical approaches. Key data management tasks will substantially affect the resultant analysis and care needs to be taken with how cues are elicited and coded prior to statistical assessment. In addition, the analysis of such data for evaluating (and in the presence of) response shift raises particular difficulties. Different forms of response shift identified in chapter 8, and case scenario examples from the study are used to illustrate these challenges. Such challenges reinforce the value of an individualised approach to assessing change and response shift. Nevertheless, some recommendations for conducting quantitative analysis using SEIQoL data are made. Recommendations include greater attention to and transparency about the categorisation of elicited SEIQoL data prior to quantitative analysis.

Preface to chapter 12

The empirical chapters of the response shift theme (chapters 6 to 10) and the current review chapter of group level assessment have addressed quality of life and response shift as their primary focus. The final empirical chapter of the thesis (chapter 12) investigates the consequences for patients of hospital waits for diagnosis and treatment for their knee problem. This was not an original aim of the thesis study, but was a pressing issue that emerged through the course of the patient interviews. It was decided to expand the scope of the interview and analysis to address this issue. In doing so this has provided a strong clinical perspective to complement the methodological focus of the other chapters of this theme.

Chapter 12: Compounding the problem, confounding the outcome - waiting for knee care in South Wales

Abstract: *Patient accounts of lengthy waits for clinical care for their knee injury predictably arose within the interview narrative. The iterative nature of the qualitative study approach allowed for patients to be questioned further about this within subsequent interviews. The exploration of these accounts was, therefore, added to the overall study analysis plan, as a secondary aim. This sub-study is presented within this chapter. Patients in the current study reported both lengthy waits for a diagnosis and for effective intervention. Patients perceived inadequacies within the NHS system, including deficiencies in the provision of information. There was patient awareness of a formal parallel referral route resulting in inequitable access to care. The additional burden of lengthy waits had emotional consequences for patients and led to fears of an increased risk of damage to their knee joints. A variety of coping responses emerged from their accounts, including practical approaches such as accessing private healthcare and increased use of analgesia. Emotional and cognitive coping responses included the use of downward social comparison. The role and adequacy of clinicians as patient advocate was also questioned. The implications of extended waits for effective clinical care are discussed. A particularly relevant consequence in the context of this thesis is that patients may adapt and undergo a response shift in their self-reported quality of life. This may benefit patients if the process helps retain quality of life levels. However, if clinical prioritisation is informed by such self-reported outcomes, failing to take account of response shift may prejudice patients who were successfully adapting to their knee injury.*

12.1 Introduction

The thesis has addressed key methodological issues (responsiveness and response shift) in the assessment of change in quality of life. The clinical context of the response shift theme involved patients awaiting and undergoing

diagnostic and therapeutic intervention for knee problems. The nature of quality of life as expressed by this sample of patients was presented in chapter 7 and served as baseline for consideration of response shift changes. Furthermore, it provided an insight into the diverse and complex way in which people perceived this construct. Studying patients at different stages of clinical management for their knee problem also provided an insight into the challenges and consequences for those often waiting long periods of time for effective intervention. Although assessing such impacts was not a primary focus of the interview study, it nevertheless was a salient outcome of the process. Therefore, the current analysis has been included to report these observations and also to provide a stronger clinical perspective with which to complement the methodological chapters of the thesis.

The effectiveness of prompt surgical intervention: Systematic reviews of surgical interventions for meniscal and posterior cruciate ligament injuries have largely failed to demonstrate conclusive evidence for their effectiveness, compared to conservative treatments.^{287 288} However, the absence of proof of benefit is not the same as proof of no benefit, and the reviews conclusions are primarily due to the absence of high quality randomised controlled trials. The same is true for interventions for anterior cruciate ligament injuries.²⁸⁹ In a survey of North American surgeons, Marx and colleagues concluded that variation in clinical opinion regarding treatment for ACL injury may reflect the relative paucity of evidence in certain areas of clinical practice.²⁹⁰ However, a systematic review by Linko and colleagues did find some advantage of surgery over conservative treatment for complete ACL rupture in terms of higher levels of knee stability and functioning (as assessed using a laxity tester and Lysholm score of functional impairment).¹⁷⁶ Similarly, Dunn and colleagues reported a retrospective cohort study of 6576 US army personnel who had been initially hospitalised for at least a knee arthroscopy, with or without ACL reconstruction.²⁹¹ The risk of subsequent knee surgery was significantly lower for those undergoing an initial ACL reconstruction (4.90/100 person years) compared to patients treated conservatively (13.86/100 person years).

In clinical practice, surgical treatment is recommended for most meniscal tears, aside from those causing only minor symptoms in less active patients.²⁹² Indeed, non-operative treatment is usually considered inadequate for patients with physical jobs and physically active lifestyles, where activity reduction may have serious consequences. Rath and Richmond reviewed the long-term success of meniscal repair and reported rates between 67% and 92% at two year follow-up, although they noted that this variance depended upon type and location of tear, and form of outcome measurement. However, one of the factors found to favourably influence healing was a short interval between injury and repair.²⁹² For example, Venkatachalan and colleagues retrospectively reviewed meniscal repairs in 62 patients.²⁹³ They found that repair within three months of injury resulted in better outcome compared to later arthroscopy (success rates of 91.6% and 58.3% respectively, based on clinical assessment).

Furthermore, for patients with anterior cruciate ligament (ACL) deficiency, delays in reconstructive surgery may increase the risk of further meniscal tears. De Roeck and Lang-Stevenson found that 10.3% of 68 UK patients waiting for ACL reconstruction sustained a meniscal tear during the wait (mean duration of 13.2 months) between initial arthroscopy and ligament reconstruction.²⁹⁴ Similarly, O'Connor and colleagues reported a large retrospective case series study of 1375 patients.²⁹⁵ They found that ACL reconstruction conducted more than six months after initial injury increased the risk of meniscal injury for males (odds ratio: 1.5) and females (odds ratio: 3.4) when compared to surgery conducted within two weeks of injury. Thus, early ACL reconstruction is recommended for groups such as manual labourers and active athletes to preserve menisci.²⁹² However, conclusive trial evidence comparing patients randomised either to immediate treatment or waiting list conditions was not found.

Local waits for treatment: In the present study, waits following either referral or being listed for arthroscopy allowed exploration of response shift. In February 2005, waiting times for arthroscopy procedures were reviewed for the three local clinical collaborators on the DAMASK study (all based within

the same clinical department).²⁹⁶ Waiting times for non-urgent patients for a routine out-patient appointment were as long as 120 weeks (mean waiting times were not provided). The longest likely wait for arthroscopic treatment (either as a day case or in-patient) following clinic consultation was 68 and 78 weeks respectively. Whilst over the course of the DAMASK study maximum waiting times increased, even at its outset the longest combined wait for arthroscopy was four years (excluding any interim period between first out-patient appointment and decision to admit).

In the context of lengthy orthopaedic waits, primary care access to MRI may allow more appropriate routing of referrals; may help expedite planned or current referrals; and may facilitate more appropriate clinical management.¹⁸¹ Within the single NHS Trust where the clinical collaborators (both specialists in Orthopaedics and Radiology) were based, MRI has also been used to screen patients on arthroscopy waiting lists following clinical diagnosis of meniscal tear and shown potential for reducing waiting list size.²⁹⁷ However, in early 2005 waits for GP requested MRI of the knee were as long as 2 years.²⁹⁸ A diagnosis may be obtained sooner via direct radiological investigation than through clinical assessment and investigation in a hospital orthopaedic clinic. However, this still represents a significant wait for patients, even before an orthopaedic consultation is achieved.

Identifying a thesis sub-study: Given lengthy local waits for radiological investigation with MRI and arthroscopy; the consequent impact on patient well-being; and the importance of timely intervention, it was unsurprising that health service delays featured as an unprompted output of the research interviews. The interview's semi-structured schedule already incorporated a narrative summary of patient experience in relation to their knee. Whilst descriptions of such waits by patients were explored further in the interview (consistent with the general iterative qualitative method), the original plan for data collection was largely unmodified. Addressing waits for diagnosis and treatment was subsequently added to the overall analysis plan as response shift could potentially have major implications for policymakers and individual

patients in relation to waiting lists. As such it formed the basis of this thesis chapter.

12.2 Methods

The aim of the sub-study was to explore patient accounts of extended waits for care in relation to their knee complaint. This could include waits at various stages of clinical management, for example, for diagnosis or surgical intervention. Descriptions of the impact of the knee injury inevitably implied an interval between onset and effective treatment. However, analytic focus was upon management delays that patients considered excessive.

Qualitative analysis thus proceeded as described in chapter five, with initial description moving through to conceptual ordering. Data from interviews at both baseline and follow-up, and from all sampled patients, were analysed. Illustrative extracts were selected for each emerging theme and presented in the text.

12.3 Results

System processes: Considerations about waiting for care included views about the National Health Service (i.e. healthcare system) processes within which patients were engaged. Not all comments were negative or indicated disadvantage, and there was also some reticence about being critical of the health service. A lengthy treatment delay was even viewed positively by one patient who had been treated previously and was wary about further intervention (II#23).

However, waits were typically the source of discontent, especially if due to perceived error or unjustifiable delay. A few patients reported feeling effectively 'lost' within the system, for example, a GP referral apparently not being received in secondary care (I#7); or not being placed upon surgical waiting lists (II#1). Patient #12 described her frustration about being passed

around within a system with associated waits at each stop. Sometimes patients made assumptions about the progress of their clinical management (for example, about being placed on a surgical list, or about the speed of the referral process) which they subsequently learned was incorrect. A few patients mentioned having appointments cancelled or not being realised, without explanation and occasionally with little notice:

Spouse of I#34: It's been cancelled. This is the third appointment ... he sort of got, came home from the little job that he were doing and they cancelled it about an hour before he was due to go up there.

(Male / 46 years old / referred / water mains manual worker)

and :

I#17: I actually saw the consultant and he said we'd have you in by the Christmas and it didn't come about.

MR: Mm

I#17: So they said, "Oh, you'll be in before the summer holidays." That didn't come about.

(Female / 51 years old / listed / school support worker)

An absence of effective information was partly responsible for assumptions made, and unrealistic expectations held by patients. For many newly referred patients followed up after six months, the only formal NHS contact they had received was a letter enquiring whether they wished to remain on the waiting list for the orthopaedic out-patient clinic. One patient, who had been initially interviewed just after referral by his GP, reflected at the second interview that he had been given no indication about how long he may have to wait for an out-patient appointment either by his GP or by the hospital. Consequently, he had initially requested only short periods of official sick leave from his GP, hoping that he would be seen and receive intervention:

I#10: I was going every month - first of all it was I was going every two weeks ... I didn't want to sign off for longer than two weeks at any one time. I was thinking, "I'm going to get a letter from the hospital".

(Male / 51 years old / referred / retail worker)

A second patient referred to the communication she had received:

II#15: Um and now I've got a letter saying that I've been assigned to a consultant in Llandough and I haven't heard anything else [laughs]
(Female / 19 years old / referred / medical student)

A common theme was concern about losing one's place upon a surgical waiting list and having to start again at the bottom. Some patients felt that their condition was either intermittently bothersome or had naturally improved, or were at least temporarily wary of undergoing arthroscopy. Foregoing an operation and subsequently having to start from scratch if their problems re-emerged was of concern to them and in some cases caused patients to pursue referral despite improved symptoms (e.g. Patient #11). As such, there was a feeling that the system was inflexible and unfair, especially for patients who were concerned about not being an unnecessary drain upon the NHS. The following comments reflect this:

II#17: .. well shortly afterwards [after declining an arthroscopy] it started playing up and I thought I could just go back and have it done. Well it doesn't work that way, does it, the National Health? So it meant that I had to go back on the waiting list and then we said, well, "I can't carry on like this. I'm in absolute agony".
(Female / 51 years old / listed / school support worker)

II#17: I felt it was unfair that I'd been pushed back on the waiting list – basically to the bottom to start again because I was trying to save everybody any hassle with it.
(Female / 51 years old / listed / school support worker)

II#18: Yes, I'm waiting to hear from a specialist. So what do I do when I have a letter to say go and see him? Do I phone them up and say, "Look I'm not having any problems with me knee"? If I do that and it comes back I'll be – and I'll go back to the bottom of the list. On the bottom of the list.
(Female / 64 years old / referred / retired)

Parallel referral processes: One feature of local NHS provision that became apparent through the interviews was a specialist knee injury clinic to which patients could self-refer. One of the referred patients (patient #9) described the specialist clinic in his account. Although he was the only patient to

describe this specialist clinic in detail (and therefore, was not typical of all referred patients), it is described below as it represented an alternative referral route for patients. His account and that of one other referred patient (patient #10) were compared to illustrate these parallel routes (table 12.3.1).

Table 12.3.1 Comparing parallel care processes within the local NHS system

	Patient #9	Patient #10
<i>Management status</i>	Arthroscopy	Waiting list
<i>Employment Impact</i>	Self-employed Not painful, "know it's there", still playing squash	Manual retail Unable to work, unable to walk distances, financial concerns, +++pain
<i>Care pathway</i>	Acute Knee Injury Clinic	GP referral
<i>Progress</i>	10 day wait for MRI	After six months, OP letter enquiring whether he wished to remain on the waiting list

Patient #9 had received his operation approximately a year after presenting to the self-referral clinic, and prior to that had waited approximately ten days for MR imaging. In the meantime, his knee problem hardly limited his daily activities – for example, he had continued playing squash:

#9: I sort of push off sometimes and sort of twists and it clicks and it's not painful but I just know it's there and that is annoying and frustrating. The same as the squash - I know it's there and I know... I've now got to the point I know, you know, I've got to put the ball down me left side and I've got to come off me left side I can't get the ball. That's it end of story.
(Male / 51 years old / listed / self-employed businessman)

He explained that the short wait for treatment was due him avoiding a GP referral. He had been led to believe that such a referral would have been accorded much less weight by staff in secondary care, than a similar request from a hospital consultant:

#9: So they [staff within radiology] explained it all to me and he said that if you had gone via your GP you would have been at least a year waiting. I said, "Why is that then?" He said, "Oh, I've got 2600 people or something

waiting ... because you're referred by a consultant in the hospital there's something definitely wrong with you – you're not swinging the lead ... unfortunately with the GP, often the GP would say I'll send you for a scan - you can use this in your research if you like – I'll send you for a scan to get them out of their hair". And he says, "so if they come with a GP referral they get to the bottom of the pile but if they come with the consultant's referral via the Acute Knee Injury Clinic or something similar, they go to the top pile", – so that's interesting.

(Male / 51 years old / listed / self-employed businessman)

Not only did the patient believe he received a scan sooner than he otherwise would have done, he also was listed for his operation much sooner than the current waiting times for arthroscopy would have indicated. This patient had been advised to refer himself to the knee clinic by a general practitioner within his own family doctor's practice. He provided an interesting contrast to the experience of the second patient (#10) who was of the same age, gender and locality. Patient #10 was interviewed following referral for a knee problem that was preventing him work in his manual job. By the second interview, he continued to be off work and had now moved in with his elderly parents to care for both them and his chronically ill wife. Unaware of how long he may have to wait, he was desperate to get back to work and return his life to normality. Although he felt supported by his GP, his doctor also appeared bewildered by the lack of apparent progress with the referral. Whilst his GP was prepared to sign him off sick for longer, this didn't address the patient's fundamental need to a return to work:

Il#10: I did have a word with the doctor on Monday and I said, "Look, I'm getting so frustrated basically that if I feel I can do it I'll get back to work. He said, "Just come back and see me if you feel that way," he said, "I'll sign you off."

...

Il#10: I just want to get back into work and, you know, get my life back on line again like I used to.

...

Il#10: ... and the one doctor [GP] there said he can't understand why this man hasn't taken this on. He said, "You're still able to work".

(Male / 51 years old / referred / retail worker)

Perceived impact of waiting for care: The impact of the index knee problem could be substantial and was viewed as affecting many domains of life.

However, the additional burden of waiting for either diagnosis or treatment was expressed by several patients. Obtaining confirmation of diagnosis from a specialist surgeon or radiologist could assuage general uncertainty about the cause of the knee problem, and also exclude concerns about more malignant pathology:

I#12: ... The waiting, I think, is the worse thing. I was so relieved when I finally seen someone and they told me what the problem was because up 'til then I though it was in my mind.
(Female / 29 years old / listed / customer relations officer)

and:

I#30: Until I knew what was wrong with it, I had all sorts of things. I had cancer and all. There were so many x-rays and MRI – it took so long.
(Female / 44 years old / listed / residential care team leader)

A few patients summarised the overall impact of waiting upon them, including the notion that life had been put on hold or curtailed until further action was taken. One patient also remarked upon the impact on her partner of undue delays. Emotional reactions to delays in the system included anxiety, frustration, irritation and annoyance. Some examples included:

I#31: It's a worry and I can't see it improving without something happening
(Male / 57 years old / referred / newspaper editor)

I#5: It's the same every day and it's sort of curtailed everything and to wait three years – well it's been horrendous to be honest
(Male / 52 years old / listed / electrician)

I#12: It's just the waiting, the waiting is horrendous.

MR: Yes

I#12: Like two and a half years just to wait for a scan
(Female / 29 years old / listed / customer relations officer)

Several patients had already described their expectation that their knee problem may render them prone to difficulties in the future with arthritis. Some patients expressed concerns that their actual wait would increase their risk of arthritis in the long-term, or lead to deterioration in their current condition in

the short-term. A related concern was the development of, and entrenchment within, a disabled role whilst waiting for effective intervention. The following extract illustrates concerns of waiting-induced change:

I#29: I think, "have I done any more damage by carrying on working".

MR: Right

I#29: By standing, by doing this and doing that. I don't know if I've done any more damage -

MR: Right

I#29: - by waiting.

(Female / 54 years old / listed / catering worker)

Response and coping processes: Facing uncertainty and delay for treatment prompted a variety of responses amongst patients. A number considered private healthcare, and a few chose it. Patient #29, for example, was asked by her GP whether she would consider a private consultation. She subsequently chose to be referred privately and was seen the same week. Influencing her decision was her concern that she was exacerbating the damage in her knee whilst it remained untreated. She wasn't happy that she felt obliged to go down this path. Patient #13 reflected on the disparity in access between private and public health care – in this instance an ongoing wait of ten months for an NHS consultation compared to three days with the private health insurance company, BUPA. Patient #30 described how she paid for a private consultation due to severe pain and mobility problems; a substantial increase in her weight; and fears about malignant pathology in the knee. The following extract illustrates these points:

I#29: Well I took more painkillers than normal and then the doctor - she asked me then eventually if I could pay privately, what would I like to do. Then it still took two years, so I wasn't amused with that, you know.

(Female / 54 years old / listed / catering worker)

I#30: I think the amount of pain I was getting, and I was still with the GP and he reckoned that the osteoarthritis which showed up on the scan wouldn't be enough to give me the exact pain – it must be something more, and he said the only way to find out was with an MRI scan. I think I waited six months for that, but in between I went to BUPA. I was getting to a stage where I really couldn't ... drive into work.

(Female / 44 years old / listed / residential care team leader)

Some patients described the role played by their own GP. As described above, one had suggested an alternative route to care via an acute knee injury clinic, which was successful in expediting care. Patient #35 described how she was concurrently referred by her GP to the orthopaedics out-patient clinic and for imaging. The patient also sensed that her GP was unable to effect the situation significantly, and that this had been a reason not to re-consult more frequently:

MR: And you haven't been back to your own general practitioner then about your knee?

II#35: I really don't see any point in it. I mean a GP's got enough to do without going back and complaining about something that basically they can't do anything about.

(Female / 63 years old / referred / retired social worker)

and:

II#7: No, I haven't bothered the GP. What can the GP do, you know ...

(Male / 53 years old / referred / local government administrator)

The situation was so bad for patient #27 that she wrote to her Member of Parliament but was merely advised to change her doctor. Other active responses by patients included taking analgesia more frequently (patient #29) and attempting to return to physical activities in spite of advice to cease:

I#19: I did completely lay off all sport then ... but then when I came back in the September - the start of the new term, I kind of gradually tried to ease myself back in because ... things were getting no further with the doctor and I just thought, well, you know, maybe it's just something I've got to get on with, maybe it'll just go.

(Male / 24 years old / listed / medical student)

Despite waiting for care, patient #28 derived consolation once she finally had a date for intervention and as such she had an identified end in sight. More passive responses to the situation included some form of accommodation or

resignation to their fate – a sense by which they had to make the best of the situation or simply accept it. For example:

II#22: The pain is constantly there all the time, whereas I've learnt to live with that and I'm not expecting to be seen for another year or two, so you've just got to get on and do the best you can.

(Male / 34 years old / referred / self-employed carpenter)

II#7: No, it's to the back of my mind - if it takes two years, if it takes three years ... the amount of time is a nuisance, you know, but it's not as if it's stopping me work or anything like that. I'm not a manual worker.

(Male / 53 years old / referred / local government administrator)

Another form of coping response was comparison with other people in similar, but worse, situations, (i.e. downward social comparison):

I#12: I just feel so sorry for some of the older people, because they're treated horrendous I think. Not too bad for me, I'm still young and I can get around but some of them have been waiting so long. It's awful - the waiting is the worst thing.

(Female / 29 years old / listed / customer relations officer)

Another comparison patients made was between the waiting times locally and in other parts of the country. Thus, patient #31 considered himself lucky to be living in South Wales and not another part of the UK (specifically mentioning London and Birmingham) where he believed the waiting times to be greater. Some patients, however, had not coped at all well with their experiences. The impact of the knee problem had been aggravated by the perception of how the system has failed them. This in turn had turned to bitterness, for example:

I#5: Well there's not a lot you can do sitting in a chair physically, I think, once you lose the physical side of your life. I think it drags you down mentally and in turn you become caustic about things – you definitely become angry, you know? "Why me? Why couldn't someone else have this?" I mean it's only my knees, and I know there's a lot of people worse off, but I don't know what they are going through so I can only relate what I'm going through - and it's not a nice feeling definitely... it does bring you down, and that's why I'm bitter with the NHS because of the time you have to wait. I feel sorry for the old people who are waiting for their hip replacements and knee replacements, but at the moment

I'm just feeling sorry for myself - so they will have to wait for my sympathy, I'm afraid.

(Male / 52 years old / listed / electrician)

The role of the clinician: There was both praise and criticism for clinicians. Several patients felt that their GP had been supportive, albeit sometimes ineffective in facilitating rapid care. The first patient described below and his doctor, seemed to be working towards different goals. The GP had encouraged him to attend less frequently and be signed off work for longer periods of time. The patient was (unrealistically) optimistic about being seen in the orthopaedic out-patient clinic and was reluctant to be signed off sick for long. Nevertheless the patient was positive about his GP. In contrast, a second patient felt that his own doctor was simply attempting to 'fob him off'.

Some extracts:

Il#10: I did mention it to him [lack of an appointment]. I said, "I'm sure there's other people a lot worse off than me". He said, "Don't make excuses for the National Health"

...

MR: Sound like your GP's quite supportive then?

Il#10: Oh, yeah

(Male / 51 years old / referred / retail worker)

Il#7: ... but you think, "alright, anything to keep you happy", you know, "and go away", which is what they seem to want, you know

(Male / 53 years old / referred / local government administrator)

Some patients felt that they had received insufficient consideration. For example, one woman felt that responsibility for her well-being was being passed around within the system and ultimately she felt abandoned and had to get on with things herself:

Il#1: ... but I had another appointment and my husband turned around and said to them ... "what do you intend to do in the meantime for her pain", ... "Oh", he said, "that's nothing to do with us", he said, "that's your own doctor and the pain killers".

...

Il#1: So basically their attitude was – "we can't do anything yet. You'll have to wait until whatever and you've just got to cope with it".

...

II#1: But it's just the pain. You know, I just feel like climbing the wall some days because it is so intense. It just seems that there's no one out there that's willing to do anything for me.

...

II#1: ... I can see myself in a wheelchair because it is getting that bad.
(Female / 41 years old / listed / retail manager)

There was a feeling amongst some patients that insufficient consideration was being given to the general impact of their knee problem; that they are not being taken seriously; and that they were generally being 'fobbed off' with references to general system difficulties:

I#33: I just wish that you got treated like, you know – “you've got a bit of a knee problem” and you know, you're just thrown to the back of the queue

...

I#33: ...if I was an office worker I'd be fine cos – but I'm not, I'm struggling ...
(Male / 38 years old / referred / builder)

and:

II#24: It could be another three months. You say, “When will it be?” and they say, “Well, it's a system”.
(Male / 56 years old / listed / glazier)

12.4 Discussion

Overview: Within this sample of patients there was evidence of physical, functional and psychological morbidity at least partly induced by waits for diagnosis and intervention. Such consequences may be viewed as iatrogenic in that they are related to the running of the medical system. Patient expectations for treatment and diagnosis are raised (for example, by an initial referral), but these expectations are not promptly met. Consequently, patients perceive an additional impact of waits which may in turn engender a sense of despair, bitterness or inequity. These findings are serendipitous in that it was not an aim of the original data collection or analysis plan, and patients were not systematically questioned about perceived delays. Indeed, study patients had merely been selected on the basis of recent referral or imminence of

arthroscopy, rather than on the grounds of recognised delay. Nevertheless, patients were probed further about such issues when they arose, commensurate with a broadly inductive and iterative approach. Furthermore, the frequency and range of largely spontaneously expressed opinions (for example, 18 of the 39 patients are quoted in the chapter) serve to emphasise the importance of this issue for patients. Whilst there is evidence of accommodation and adaptation which may serve to ameliorate quality of life, this may still mask the true burden for such patients who experience lengthy delays in effective physical therapy.

Social and clinical equity: Whilst waits experienced by patients locally for arthroscopy and MRI may be longer than in other areas of the UK, there were also disparities apparent for patients being cared for within the local health service. In both cases, this raises the issue of equity of access within a publicly funded health service. Patients with more urgent conditions should receive services ahead of those with less urgent conditions, and those with the same degree of urgency should wait the same time regardless of geographical location.^{241 299}

In their study of Canadian patients awaiting major joint arthroplasty, Kelly and colleagues distinguished between social and clinical equity.³⁰⁰ Only 10% of observed variance in waiting time could be explained by modelled factors including socio-demographic, clinical, health system and health status. Whilst they concluded that access was equitable in terms of social factors, waiting times were managed unfairly as far as clinical equity was concerned. From our study data one cannot draw such conclusions about social equity. Nevertheless, the divergent experiences of the two patients described in table 12.3.1, and the difference in waits of all local patients compared to other UK regions points to substantial clinical inequity.

The acceptance of long waits: In a study of patients waiting for hip and knee replacement Llewellyn-Thomas and colleagues found that lower tolerance for waiting was associated with lower reported utility scores and shorter waiting times since surgery was recommended.³⁰¹ Drummond and colleagues found

that amongst patients listed for cataract surgery in a cross-cultural study in Manitoba, Barcelona and Denmark, the acceptability of perceived wait was most strongly predicted by estimated personal wait and level of impairment.³⁰² Socio-demographic factors did not predict tolerance but coping abilities and quality of life variables were not assessed in the study. Tolerance may also vary by site-specific factors, such as awareness of referral and waiting processes locally and perceived appropriateness of surgery.^{302 303} In our study, as in others, it is clear that information about process, including expected waits and decision process, is sometimes poorly communicated to patients, from both primary and secondary care.^{303 304} The appropriateness of leaving patients incompletely informed about very lengthy waits for diagnosis and treatment, whilst they make decisions about their work and general functioning, is questionable.

Prioritising access to healthcare: Waiting times for patients do not necessarily reflect observed levels of pain or disability, although a confusing factor in assessment is variable use of terms.^{263 300 305 306} Internationally there have been attempts to prioritise patients waiting for secondary care on the basis of explicit criteria, rather than implicit clinical judgements.^{306 307} The New Zealand Priority Criteria Project aimed to address the problem of lengthy waiting lists for elective surgery with a move towards specific booking times.³⁰⁷ Derrett and colleagues assessed patient perceptions of the Clinical Priority Assessment Criteria (CPAC) used in New Zealand and the resulting short notice periods for surgery were found to be generally acceptable.³⁰⁸ However, patients were anxious that if they were unable to meet the appointed date for elective surgery they would fall to the bottom of the list. Furthermore, 'not currently eligible' patients reported insufficient communication and frustration regarding incorrect information when it was received - problems also evidenced in our study. Derrett and colleagues concluded that including quality of life assessment in clinical practice would improve care by informing the prioritisation process, but also cautioned that increasing the use of patient questionnaires may induce gaming. That is, patients may exaggerate their symptoms to get surgery sooner. In Wales, Edwards and colleagues found broad support from professionals and the

public alike for an explicit priority scoring system to determine waiting times for elective surgery.³⁰⁹

Measuring the impact of lengthy waits: What harms result for patients facing lengthy waits for elective orthopaedic surgery? There is evidence for the benefit of prompt treatment for meniscal tears and anterior cruciate ligament reconstruction for physical well-being.^{292 294} It is clear that patients in the present study themselves perceive both physical, psychological and social harm as a consequence of waiting for surgery. Indeed, the detrimental consequences of waiting have been observed, for example, by Ostendorf and colleagues who reported a reduction in quality of life in patients whilst waiting for hip arthroplasty.³¹⁰ However, Derrett and colleagues found no association between length of wait and levels of adverse health status in a survey of patients waiting for either prostatectomy or elective knee and hip arthroplasty in New Zealand.³⁰⁴ Two studies of patients with osteoarthritic joints by Peters and colleagues, and by Mahon and colleagues, found that self-reported health status and functioning did not inevitably deteriorate over time, and may actually improve prior to intervention.^{311 312}

Response shift as an additional patient burden: Kelly and colleagues attributed apparent stability in reported health status amongst patients awaiting joint arthroplasty to symptom exaggeration and minimisation at different time-points, and a possible ceiling effect.³¹³ Similarly, the positive correlation between tolerance of waiting time and length of wait led Llewellyn-Thomas and colleagues to conjecture that such 'stoic' patients may have adapted to their living conditions over time.³⁰¹ It is possible that patients facing lengthy waits for physical intervention may well have adapted consciously or sub-consciously to their predicament. Thus, self-evaluation of health status may reflect that process, and show stability or even improvement. Lewis and colleagues have suggested that time on waiting lists is advantageous in allowing time to reflect on treatment decisions and adaptation to a new condition.²⁹⁹ However, if such adaptation is accompanied by response shift, continuing physical and functional burden may be masked. Using health status and quality of life measures to prioritise surgical waits without

considering the possibility of response shift may actually penalise patients who have already adapted to their circumstances whilst waiting. This is true even if such patients report general life satisfaction, for they may still have physical health deficit which is potentially improvable.

12.5 Summary

Patients included in the Response Shift study faced hospital waits for both diagnosis and treatment of their knee problem. The nature of such waits and its impact upon well-being was an emergent aspect of the interviews for those in both waiting list and arthroscopy groups. Topics raised by interview respondents included descriptions of health service systems, alternative referral pathways, perceived impact of waiting for care, individual coping and response processes, and the role of the clinician. An overarching theme was the adequacy of information provided to patients both at a systemic and individual clinician level. Given the challenges of substantially reducing imaging and clinical waiting times, there may nevertheless be some scope for addressing how patients are informed about both their condition and its management. In the context of response shift there was further reflection about the potential disadvantage for patients who have adapted psychologically whilst waiting for care. Finally, the chapter as a whole provides a clinical perspective to complement the methodological emphasis of the other chapters of this thesis.

Preface to chapter 13

The current chapter concludes the empirical work of the response shift theme of the thesis. Within each chapter results and discussion have been presented enabling each to be read as a largely autonomous unit (albeit with themes connecting each chapter). The following and final chapter synthesises the work presented in all the chapters of both research themes. As such it provides a summary of the thesis, addresses practical implications and limitations and makes suggestions for further work.

Chapter 13 Synthesis and summary

13.1 Introduction

This thesis has been constructed upon two methodological themes, both of which relate to the measurement of change in quality of life: *scale responsiveness* and *response shift*. The broad aim of the responsiveness theme was to describe the performance characteristics of a commonly reported effect size statistic under different distributional conditions. The broad aim of the response shift theme was to assess the presence and nature of response shift in a sample of patients at different stages of clinical management for a knee injury. As the studies within each theme have been presented in autonomous chapters, the purpose of this concluding chapter is to summarise and draw together the principal findings. For each theme, the main findings from each chapter are summarised; a commentary about their relevance and importance is provided; study limitations are highlighted and suggestions for future work are made.

13.2 Responsiveness theme

Summary of findings: Simulation studies were used to assess the performance of the standard response mean (SRM) effect size. In smaller samples, the three modelled approaches for handling negative change scores produced markedly different effect size values. Omitting or reversing negative values prior to calculation increased the value of effect size compared to the approach which used such values as generated. The latter approach was therefore used for subsequent modelling. Simulation demonstrated that the relative bias of effect size estimates was usually small and generally positive (i.e. over-estimating the true effect), especially with larger sample sizes. This was true for both parametric and non-parametric SRMs calculated from values drawn from a normal distribution.

The positive bias in calculated effect size results in an effect size slightly larger than the 'true' underlying value (as determined by the parameters of the simulation model). With data drawn from an underlying normal distribution, relative bias was never greater than 8%. However, when data was generated from log normal distributions, parametric estimates of SRM were substantially biased, especially with smaller samples and small true effect sizes.

Commentary and practical implications: Amid the uncertainty about how responsiveness should be evaluated, the performance characteristics of individual effect sizes have largely been either assumed or ignored. The present studies provide reassurance that the SRM exhibits little bias when sample size, mean underlying effect size and shape of underlying distribution is varied. The studies highlight the impact of different methods of calculating SRM (dealing with negative values; parametric versus non-parametric equivalents) upon the 'true' and therefore, derived value. Effect sizes derived using different approaches to handling negative values (and therefore the studies from which they are reported) may not be simply compared. The comparison of parametric and non-parametric SRMs must also account for the inherent difference in magnitude between the two statistics. Perhaps most importantly, parametric SRMs calculated from data drawn from a log normal distribution may provide a misleading and inflated estimate of change. In such circumstances, the non-parametric SRM provides a relatively unbiased estimate, and should be more widely applied than it currently appears to be. The results of the study are also a reminder that the distributional nature of change scores should be assessed prior to calculating an effect size. Where the distribution is not found to be normal, options may include non-parametric statistics such as that evaluated in this study and also data transformation.

Limitations: The approach chosen used a single distribution to model outcome measure change scores. The modelling program also used the zero change score as a simple proxy cut-point for classifying the 'improvement' or 'deterioration' that would be identified in empirical studies by self-reported transition items. There were two potential limitations in the modelling process worth noting. Firstly, the adequacy of a single distribution to model 'improved',

'stable' and 'deteriorated' patients could be questioned if, for example, one considered that patients in intervention and control groups constituted separate distributions. However, a single distribution is more justifiable when modelling the process of change within a single study group (for example, within a cohort study or within an intervention group). It is also an appropriate model for the particular effect size statistic (SRM) investigated. Nevertheless, alternative modelling approaches could embrace the generation and combination of distinct distributions of changed and stable cases. This would be especially important for modelling the index of responsiveness which, unlike the SRM, specifically compares 'improved' and 'stable' patients.

A second consideration was the suitability of using zero change score as a simple proxy cut-point for sub-group identification. In practice, 'changed' sub-groups formed by transition items may well include patients with change scores in the 'wrong' direction (i.e. negative change scores for improved patients). The modelling process therefore presents a simplistic model of what may occur in 'real life'. However, if effect size calculation uses all change scores as generated, this becomes less important. Furthermore, the validity of transition items themselves has also been recently questioned.

Suggestions for further work: Further modelling work is indicated by the current study, for example, of the standardised effect size and the index of responsiveness. This would require an amended approach to the modelling whereby separate underlying distributions would be used to generate baseline and follow-up cases (for SES); and 'improved', 'stable' and 'deteriorated' cases (for IoR). Calculating the three effect sizes from the same model would indicate the relative performance of each statistic under varying distributional circumstances. Although relative bias was not generally substantial in the modelling studies, if further modelling indicated that it was the case, it is possible that correction factors may be developed and applied. Investigation of the cause of the effects demonstrated in the modelling studies (for example through algebraic analysis) is also indicated.

13.3 Response shift theme

Summary of findings: Baseline assessment of quality of life content mapped out how respondents with internal derangement of the knee conceptualised quality of life. Twelve major themes emerged. The analysis raised questions about the meaning of the quality of life construct, and the extent to which the assessment method (SEIQoL-DW) measures quality of life itself, or factors which contribute to it. Mapping out quality of life domains highlighted the complexity of the construct and, of importance for this study, the challenges for assessing construct change. The individualised approach provided a balance to conventional functionally based assessment.

Whilst some stability in quality of life content and the importance attached to constituent domains was apparent across assessments, there was much credible evidence of change. The analysis distinguished between apparent and real changes; identified change that may have occurred outside the timescale of the study period; and described the role of both health and non-health related catalysts of response shift. That re-prioritisation and re-conceptualisation may be different levels of the same process is proposed. Specifically within the context of SEIQoL-based response shift assessment, cues that decrease in importance may no longer remain as one of the top five elicited cues. It may however, still be of some importance to the individual (e.g. theoretically ranked sixth). A typology of change was developed, drawing upon response shift theory and which was also empirically based. The typology drew particular attention to more subtle changes which may not be so apparent when simply comparing SEIQoL-DW cue profiles. Processes that serve to facilitate and inhibit response shift were identified and explored. In particular, processes of coping and adaptation were delineated, with the latter incorporating elements of modification, cessation and substitution.

The qualitative evaluation of response shift was supplemented by the more commonly applied 'thentest' method. Significant change in health utility was only evident when change derived from prospective (rather than retrospective)

self-report was used. There was evidence of re-calibration response shift for health utility (arthroscopy patients) and for individualised quality of life (waiting list patients). Further analysis of response shift in health utility revealed no evidence of socio-demographic or clinical correlates. There was some support for the validity of retrospective ratings of baseline status, but also equivocal evidence for the validity of the health transition items.

A secondary (combined quantitative and qualitative) analysis explored and found evidence for the association made by respondents between the level of a quality of life cue (i.e. how well things are going) and the degree of importance attached to it. The direction of this relationship could vary and there was also some evidence that the nature of this relationship changed over time.

The use of quantitative SEIQoL-DW data for the purpose of group comparison in both cross-sectional and longitudinal designs was critically reviewed. Key issues determining the validity of such comparisons were identified - most importantly the adequacy of the classification framework, and the process of coding cues. Data from the response shift interview study was used to illustrate the challenge of assessing response shift with individualised quality of life data.

Finally, evidence emerged that physical, functional and psychological morbidity was perceived by patients to be the consequence of their lengthy wait for diagnosis and treatment. There was evidence that patients facing such delays had accommodated and adapted, with the consequent potential for response shift. The role of formal provision of information to patients about the course of clinical management, and treatment efficacy was raised.

Commentary and practical implications: The baseline analysis of quality of life content provides a contextualised patient-based assessment of quality of life which extends beyond a primary focus upon functional ability. Nevertheless, it is also clear that even affected physical function is manifest in

diverse ways. The thematic analysis highlights the challenge inherent in comparing nominally similar content domains.

The study blended qualitative and individualised approaches in assessing change. The empirically grounded typology of change extends the conceptualisation of response shift changes. It thus moves beyond the simple comparison of cue profiles, an approach which may be misleading. The study provided evidence that both transient and long-term changes play a role in response shift, and that the catalyst for response shift may be due to non-health factors. Within the framework of an individualised assessment, response shift may have little impact on overall quality of life (as represented by an index score), even though important change may have occurred. Similarly, whether a change is reported as re-prioritisation or re-conceptualisation may be somewhat method dependent.

The study of re-calibration response shift was secondary to the main analysis and was primarily exploratory. It was not therefore, supported by a formal sample size calculation and may have been under-powered. However, it supported the distinction between prospective and retrospective based accounts of change. It provided evidence of response shift in patients who have undergone both invasive and no medical intervention. The study provided evidence that retrospective assessment of baseline state (thentest) and self-reporting of health transition are separate processes, with some validation of the former, but equivocal evidence for the latter. The latter is particularly important, because of the use of transition items in assessing responsiveness and test-retest reliability.

The moderate association between ratings of cue level and weight suggests that patients do not confuse the two concepts when completing SEIQoL, which supports its validity. More importantly, in the context of this study, it suggests a potential mechanism for adaptation and thus response shift. However, the secondary and exploratory nature of the analysis, together with the small sample size, means that the results are suggestive rather than conclusive.

Several authors have presented quantitative (group level) analysis of SEIQoL and such analysis may also form a part of a response shift assessment. However, using such data needs to account for the complex and idiosyncratic nature of the cue content. Established principles for classifying and coding qualitative data should be applied and reported in future studies. There are particular challenges for quantitative analysis of SEIQoL cue data when assessing response shift which in part relate to the nature of changes represented by the typology described above.

The observation of long delays for NHS intervention, and the consequent morbidity associated with it, was serendipitous but credible. It highlights clinical inequity both within the local service, and across services nationally. That patients accommodate and adapt to such delays may enable them to retain quality of life, whilst tolerating decreased function. Such changes may be reflected in response shift. However, if quality of life influences clinical prioritisation, successful adaptation may actually disadvantage patients with continuing physical problems.

Limitations: Whilst the strengths and weaknesses of the response shift studies are discussed within individual chapters, a number of limitations are worth addressing in summary. Firstly, in some instances SEIQoL cue elicitation failed to identify cues that may have been expected to be important to individuals. If this was indeed the case, it may have been due to cues being taken for granted by patients. Whilst this may be an issue with the use of individualised measures in general, in this study it could potentially confound comparison of cues over time. However, if certain areas of life are consistently overlooked this has less impact upon response shift assessment. The probes used in the follow-up interview to address inconsistent cue elicitation would also serve to enhance the validity of cue comparison. The non-provision of personally important cues could though potentially distort the emergent framework of quality of life themes. However, the total number of completed assessments and the eventual breadth of content suggest that this is unlikely to have been the case.

Secondly, the principal categorising variable in both the qualitative and quantitative analysis was stage of clinical management. Whilst it was clear that this represented a meaningful difference between the two groups of patients, it was also clear that each group contained patients at various stages and experiences of clinical management. Whilst further analysis of group differences is still possible with this data set, the purposive sample stratification has nevertheless served to maximise sample variability as intended.

Thirdly, although interview follow-up was planned to be five months after arthroscopy, some patients may still not have either fully recovered, or even benefited at all, from the intervention. Nevertheless, they were all expected to have experienced some change in their physical health and functioning, in contrast to the waiting list patients whose physical condition remained stable. Notwithstanding the within-group variety noted above, the requirement to compare changed and stable groups of patients was still broadly satisfied.

Finally, the sample size was determined by the requirements of the qualitative analysis, rather than the thentest analysis (which was itself a secondary aim of the overall response shift interview study). Consequently, although some statistically significant associations were demonstrated, the study may have lacked sufficient power to detect other relationships. Further multivariate investigation of the correlates of response shift would certainly require a larger study sample.

Implications for quality of life assessment: This work has demonstrated some of the limitations associated with assessing response shift using superficial cue profiles derived from individualised assessments. However, it does show the value of extending the individualised approach - both for understanding how quality of life is constructed by individuals and also how changes may occur over time. As such this work provides strong support for a combined individualised and qualitative approach.

The derived typology of change does not change fundamentally the response shift framework presented by Sprangers and Schwartz.¹⁰ Rather it furthers the method for revealing and classifying response shift using an individualised approach. It provides an empirically-based cautionary note about the use of individual cue profiles to infer change. This is the case not only for response shift assessment, but also any quantitative assessment of individualised quality of life data.

Attention should also be given to the possible exclusion of cues expected to have an influence upon quality of life and, in contrast, the inclusion of cues that may not. This addresses the validity of the particular individualised method. If potential cues are not being provided by respondents because they are simply taken for granted (and therefore not expected to be mentioned) this may reduce the validity of the cue profile. This could be addressed by the addition of set probes. Nevertheless, inconsistent provision across assessments may not necessarily reflect true change – and may reflect other response biases. Again the ability to probe in a qualitative interview helps to address this potential problem.

This thesis has linked apparent re-prioritisation and re-conceptualisation response shift, regarding them as different levels of the same process. Whether a change in the importance of a cue is represented as re-prioritisation or re-conceptualisation may depend upon how much cue weight reduces and the relative importance of other background cues. Distinguishing between re-prioritisation and re-conceptualisation on the basis of SEIQoL cue profile may result in a distinction which is rather method-dependent. Not limiting the number of cues may usefully remove this distinction in type of response shift – all change might simply be reduced to re-prioritisation. However, this is not to suggest that re-conceptualisation does not exist and that it is not distinct from re-prioritisation. Rather, this conclusion may reflect that SEIQoL-DW is addressing determinants of quality of life rather than being a direct measure of the construct.⁵⁹

Suggestions for further work: The relative absence of expected cue content may indicate probing for content that is individually important but may otherwise be taken for granted by respondents. This would be relevant for response shift studies, and individualised quality of life assessment more generally. Further qualitative work may usefully address how individuals aggregate cue content when forming overall judgements about their general quality of life, and whether this adequately represents the sum of the individual components.

Continued work is indicated to explore and validate the typology of change. Furthermore, the occurrence of response shift independent of health-related catalysts (e.g. through the passage of time and maturation) could be explored through longitudinal assessment. The qualitative approach to response shift assessment used in the present study based upon a framework of individualised assessment, seems a valuable vehicle for addressing both developments. Another area worth addressing would be exploring barriers to adaptation and response shift.

Future thetest studies would be enhanced by including an external criterion of change. Further quantitative investigation of the correlates of response shift should incorporate psychological variables, as well demographic and clinical variables. Given uncertainty about the validity of retrospective accounts of change, further work exploring the processes whereby respondents form such judgements is indicated. Cognitive debriefing approaches may be particularly helpful in this regard.

The relationship between rating of cue level and cue weight should be studied further. A prospective longitudinal assessment could explore how this relationship changes over time, and how it may be affected by disease trajectory. Such an analysis would determine the role other variables may play in mediating this relationship, particularly psychological and socio-demographic variables.

The role of formal information provision within the NHS to patients about their clinical management should be further investigated. Information about timescales for treatment; likelihood of treatment efficacy; and even comparison with other similar patients may affect patient expectations. This in turn may affect patient ability (or preparedness) to adapt. Much of this would be redundant if waiting times for effective diagnosis and intervention were reduced. In the mean time though, it may serve to enhance equity and enable patients to take informed decisions about their lives and lifestyle.

Clinical implications: The thesis provides enhanced understanding of how knee problems may both directly and indirectly affect patient quality of life. The expanded scope of quality of life presented (beyond a primary focus upon health status and function) indicates a need for broader clinical evaluation of the consequences of knee injuries. Knee problems may have consequences which are not limited to mobility and pain, but may influence life in other more subtle ways. The studies are also a reminder that patient quality of life is constructed by factors which extend beyond health and illness and includes psychological and particularly social drivers. This reinforces a model of healthcare which aims to understand the whole person in context within a patient-centred approach to clinical assessment.³¹⁴ In summary, whilst patient quality of life can be affected in broad and subtle ways by a knee problem, other factors may be equally or more influential to their overall self-appraisal and clinical assessment should be cognisant of this. Although individualised approaches to quality of life assessment have been developed for mostly research application, their role in facilitating patient-centred consulting continues to be explored.¹⁹⁷

The importance of effective communication and good quality information for patients throughout their clinical management was highlighted. Waiting times for investigation and treatment will vary and reducing such waits in any one locality may be difficult to achieve quickly. However, eliciting patient information needs and optimising information provision may be more immediately achievable. This may still present some challenges, especially if

patients maintain unrealistically optimistic estimates of treatment promptness and efficacy.

Understanding how individuals construct their self-evaluation of quality of life, and how response shift may occur offers some clinical pointers. Effecting response shift may become a clinical goal rather than simply a research observation.¹⁶¹ A clinical aim to effect scale re-calibration may result in improved quality of life in the absence of any physical improvement. Where effective therapies do not exist this may be a useful strategy. However, an undue focus on response shift as a clinical goal may mean that potential therapies are not explored or developed.

13.4 Responsiveness and response shift: making the connection

For the most part, responsiveness and response shift have been evaluated and reviewed independently, although they both directly determine the ability to measure change. Methods to assess responsiveness though, may be strongly influenced and challenged by the operation of response shift.^{18 315} Perhaps fundamentally, all effect size methods assume a common metric between baseline and follow-up scores. At the very least, re-calibration response shift would challenge this assumption, re-conceptualisation could render a derived change score meaningless.

Response shift may affect levels of observed scale responsiveness. Kern and Brown concluded that response shift in quality of life (compared to health status) led to a reduction in responsiveness in patients with chronic neurological problems.⁶⁰ Pouchot commented that individualised methods which allow respondents to select domains can decrease sensitivity if the chosen domains are ones in which patients can not be expected to improve.²¹⁵ If response shift is not directly measured, its impact on treatment effect will neither be recognised, nor taken into account in analysis.

However, where response shift has been detected, for example, through the use of a t-test, the retrospective pretest score has been used to calculate the effect size (e.g. Timmerman and colleagues).¹²⁷ Similarly, Fischer and colleagues found that retrospective SRM values were approximately twice the size of prospectively derived SRMs in a study of self-reported pain and disability in patients with arthritis.³¹⁶ Thus, regardless of the relative bias inherent within any one method of calculating an effect size, presence and consideration of response shift may play an important part in the conclusions to be drawn in evaluative studies. Schwartz and Rapkin argued for a reconsideration of psychometrics of quality of a life assessment in the light of response shift.⁶⁵ They suggested that scales should be recalibrated so that effect sizes are calculated only after statistically adjusting appraisal parameters. Although they provide no further method on this point, they suggest that such an approach would increase sensitivity to change.

The actions of response shift and nature of scale responsiveness fundamentally bear upon the validity and interpretability of study results.⁵⁷ For example, the absence of conventionally observed treatment effect may be explained by a true lack of treatment efficacy. Alternatively, the outcome measure may be insufficiently responsive to the presence of important clinical change. A further explanation could be that the patients have changed, but that this has been masked by the operation of response shift. Even when change has been detected, interpreting what this represents may not be straightforward, and may still be influenced by response shift changes.¹⁰⁵

13.5 Conclusion and contribution

This thesis has made an original contribution to the assessment of change in quality of life in the following ways:

- The robustness (and thus validity) of the SRM to varying distributional parameters of sample size and true effect size has been established. This thesis therefore provides support for its continued application.

- However, this work has found substantial differences in SRM values derived from different methods of dealing with negative values. This has not previously been recognised. Therefore, future comparison of effect sizes must account for the different methods used in their derivation, including variation in the overall study design.
- Furthermore, this work has found substantial bias in effect size (relative to underlying true value) for parametric SRMs when data was drawn from a skewed distribution. The implication for future responsiveness assessment is that the distributional nature of changes scores must firstly be determined and if skewed, a non-parametric SRM should be used.
- Response shift has been evaluated most often in patients with life-threatening conditions such as cancer. In contrast, this thesis adds evidence of response shift for patients with a disabling mechanical (knee) condition.
- The thesis has extended the conceptualisation of response shift change through a typology which accounts for the complexity and dynamism of individualised experience of quality of life. This typology adds to the methods of response shift evaluation and should inform future study of change phenomenon.
- An individualised quality of life measure has been successfully adapted and used concurrently with a more commonly applied test approach for response shift assessment. This adds to the developing methods for response shift evaluation.
- Although re-prioritisation and re-conceptualisation have previously been proposed as independent forms of response shift, this thesis has provided evidence that they are more closely aligned. This has therefore contributed to the theory of response shift.
- The thesis has found evidence of an interaction between the level of satisfaction with a quality of life domain and the degree of importance attached to it. This provides one mechanism by which response shift may occur and quality of life homeostasis is maintained.

- The thesis has critically appraised the quantitative use of an individualised measure of quality of life (SEIQoL). The thesis provides recommendations for future quantitative analysis of SEIQoL data for assessing response shift, in particular the requirement for rigorous and transparent classification and coding of cues.
- As a whole, the nature of response shift has been explored by multiple and complementary methods to provide a complex and contextualised picture of quality of life experience and change.

Appendices

- I Fortran (source file) program ia_na.f95 (Responsiveness modelling: Study I)
- II Fortran (source file) program ib_nav1.f95 (Responsiveness modelling: Study II)
- III Background publications from response shift literature review
- IV Patient approach letter (Radiology)
- V Patient approach letter (Orthopaedics)
- VI Patient information sheet
- VII Baseline interview schedule
- VIII Follow-up interview schedule
- IX Individual case summaries
- X Publications and presentations

Appendix I Fortran (source file) program ia_na.f95 (Responsiveness modelling:
study I)

```
PROGRAM ia_na

! Study I modelling
!
! Note: program based upon cmb_ds16a.ftn

! define program variables

! Note that array specifications for:
!
! x set to highest desired value and actual values are to be input
! later on in the program
!
! srm set to number of simulation runs (n=10000)

! variables for individual samples

INTEGER,dimension(6)::nx = (/15, 30, 45, 90, 180, 360/)
INTEGER nxpos, nxneg
INTEGER mc_val(6), mc_mv(6) !number of samples with lt 4 values
INTEGER:: nn = 6
REAL*8 x(360),xb(360), xc(360)! Individual change score values
REAL*8 xbara, xbarb, xbarc
REAL*8 sa, sb, sc
REAL*8,dimension(10000)::srma, srmb, srmc !sample resp. statistic
REAL*8 mux
REAL*8 sdx
REAL*8 es
REAL*8 biasa_p, biasb_p, biasc_p
REAL*8 xsuma, xsumb, xsump, xsumn
REAL*8 xsum2a, xsum2b, xsum2p, xsum2n
REAL*8 s2a, s2b, s2c
REAL*8 s2ia, s2ib, s2ic

INTEGER n_sim !number of simulations being run

! sample non-parametric statistics variables

REAL*8,dimension(10000)::pqv1a, pqv1b, pqv1c
REAL*8,dimension(10000)::pqv3a, pqv3b, pqv3c
INTEGER,dimension(10000)::pqv1ia, pqv1ib, pqv1ic
INTEGER,dimension(10000)::pqv3ia, pqv3ib, pqv3ic
REAL*8 qv1a, qv1b, qv1c
REAL*8 qv3a, qv3b, qv3c
REAL*8 iqr_va, iqr_vb, iqr_vc

INTEGER,dimension(10000)::va, vb, vc
REAL*8,dimension(10000)::Pva, Pvb, Pvc
REAL*8 va_med, vb_med, vc_med
```

Appendix I Fortran (source file) program ia_na.f95 (Responsiveness modelling: study I)

```
REAL*8,dimension(10000)::srma_np, srmb_np, srnc_np
```

```
!! simulation summary variables
```

```
! variables for calculating sd
```

```
REAL*8 esa_sum, esb_sum, esc_sum  
REAL*8 ena_sum, enb_sum, enc_sum  
REAL*8 esa_sum2, esb_sum2, esc_sum2  
REAL*8 ena_sum2, enb_sum2, enc_sum2  
REAL*8 esa_bar, esb_bar, esc_bar  
REAL*8 ena_bar, enb_bar, enc_bar  
REAL*8 esav, esbv, escv  
REAL*8 enav, enbv, encv  
REAL*8 esasd, esbsd, escsd  
REAL*8 enasd, enbsd, encsd
```

```
! variables for calculating srm median
```

```
REAL*8,dimension(10000)::Ya, Yb, Yc  
REAL*8,dimension(10000)::Ya_np, Yb_np, Yc_np  
INTEGER,dimension(10000)::Aa, Ab, Ac  
INTEGER,dimension(10000)::Aa_np, Ab_np, Ac_np  
REAL*8 esamd, esbmd, escmd  
REAL*8 enamd, enbmd, encmd
```

```
! variables for calculating srm interquartile range
```

```
REAL*8 espq1, espq3  
REAL*8 esaq1, esaq3, esa_iqr  
REAL*8 esbq1, esbq3, esb_iqr  
REAL*8 escq1, escq3, esc_iqr  
REAL*8 enpq1, enpq3  
REAL*8 enaq1, enaq3, ena_iqr  
REAL*8 enbq1, enbq3, enb_iqr  
REAL*8 encq1, encq3, enc_iqr
```

```
INTEGER espq1i, espq3i  
INTEGER enpq1i, enpq3i
```

```
! variables for calculating non-parametric bias estimate
```

```
REAL*8 lqrt_np, uqrt_np, iqr_np  
REAL*8 es_np  
REAL*8 biasa_np, biasb_np, biasc_np
```

```
! set random seed
```

```
CALL DATE_TIME_SEED@
```

```
! specify number of simulations and sample parameters
```

```
n_sim = 10000
```

Appendix I Fortran (source file) program ia_na.f95 (Responsiveness modelling: study I)

```
mux = 1
sdx = 10
```

```
! calculate para/non-para effect sizes
es = mux/sdx
```

```
lqrt_np = mux - (sdx*0.67449)
uqrt_np = mux + (sdx*0.67449)
iqr_np = uqrt_np - lqrt_np
es_np = mux / iqr_np
```

```
! Start simulation
```

```
PRINT*, 'Study I(a)'
PRINT*, "
PRINT*, 'Program ia_na'
PRINT*, "
PRINT*, 'This simulation uses:'
PRINT*, "
PRINT "(a,2X,f6.4)",' - para effect size of ', es
PRINT "(a,2X,f10.8)",' - non-para effect size of, es_np
```

```
Do m = 1, nm
```

```
esa_sum = 0
esb_sum = 0
esc_sum = 0
ena_sum = 0
enb_sum = 0
enc_sum = 0
esa_sum2 = 0
esb_sum2 = 0
esc_sum2 = 0
ena_sum2 = 0
enb_sum2 = 0
enc_sum2 = 0
mc_mv(m) = 0
```

```
DO k = 1, n_sim
```

```
xsuma = 0
xsumb = 0
xsump = 0
xsumn = 0
xsum2a = 0
xsum2b = 0
xsum2p = 0
xsum2n = 0
xbara = 0
xbarb = 0
```


Appendix I Fortran (source file) program ia_na.f95 (Responsiveness modelling: study I)

```
s2ia = 0
s2ib = 0
s2ic = 0
s2a = 0
s2b = 0
s2c = 0
sa = 0
sb = 0
sc = 0
nxpos = 0
nxneg = 0
va_med = 0
vb_med = 0
vc_med = 0
```

! Generate change values

```
DO i = 1, nx(m)
  x(i) = random_normal(mux, sdx)
  xb(i) = x(i)
  if (xb(i) < 0) then
    xb(i) = (xb(i)*(-1))
  end if

  if (x(i)<0) then
    xsumn = xsumn + x(i)
    xsum2n = xsum2n + x(i)**2
    nxneg = nxneg + 1
  else
    xsump = xsump + x(i)
    xsum2p = xsum2p + x(i)**2
  endif
ENDDO
```

```
nxpos = nx(m)-nxneg
```

!calculate sample mean, sd and median for each of the three methods

!method a - all cases, as generated

```
xsuma = xsump + xsumn
xsum2a = xsum2p + xsum2n
xbara = xsuma /nx(m)
s2ia = (xsum2a - (xsuma**2)/nx(m))
s2a = s2ia/(nx(m)-1)
sa = DSQRT(s2a)
srma(k) = xbara /sa
```

```
CALL DSORT@(va,x,nx(m))
```

```
DO i = 1, nx(m)
  Pva(i) = x(va(i))
```

Appendix I Fortran (source file) program ia_na.f95 (Responsiveness modelling: study I)

```

ENDDO
IF (MOD(nx(m),2)==0) THEN
  va_med = (Pva(nx(m)/2) + Pva(nx(m)/2+1))/2.0
ELSE
  va_med = Pva(nx(m)/2+1)
END IF

! calculate position of upper and lower quartiles for v
pqv1a = ((nx(m)+1)*.25)
pqv3a = ((nx(m)+1)*.75)
! convert these values to integers
pqv1ia = int(pqv1a)
pqv3ia = int(pqv3a)
! calculate vb1 & vb3 (25th & 75th centile for v)
qv1a = Pva(pqv1ia)+(Pva(pqv1ia+1)-Pva(pqv1ia))*(pqv1a-pqv1ia)
qv3a = Pva(pqv3ia)+(Pva(pqv3ia+1)-Pva(pqv3ia))*(pqv3a-pqv3ia)
! calculate iqr for v
iqr_va = qv3a - qv1a
!calculate np srm
srma_np(k) = va_med / iqr_va

!method b - all cases, negative sign reversed
xsumb = xsump + ((xsumn)*(-1))
xsum2b = xsum2n + xsum2p
xbarb = xsumb/nx(m)
s2ib = (xsum2b - (xsumb**2)/nx(m))
s2b = s2ib/(nx(m)-1)
sb = DSQRT(s2b)
srmb(k) = xbarb/sb

CALL DSORT@(vb,xb,nx(m))
DO i = 1, nx(m)
  Pvb(i) = xb(vb(i))
!Print*, pvb(i)
ENDDO
IF (MOD(nx(m),2)==0) THEN
  vb_med = (Pvb(nx(m)/2) + Pvb(nx(m)/2+1))/2.0
ELSE
  vb_med = Pvb(nx(m)/2+1)
END IF

! calculate position of upper and lower quartiles for v
pqv1b = ((nx(m)+1)*.25)
pqv3b = ((nx(m)+1)*.75)
! convert these values to integers
pqv1ib = int(pqv1b)
pqv3ib = int(pqv3b)
! calculate vb1 & vb3 (25th & 75th centile for v)
qv1b = Pvb(pqv1ib)+(Pvb(pqv1ib+1)-Pvb(pqv1ib))*(pqv1b-pqv1ib)
qv3b = Pvb(pqv3ib)+(Pvb(pqv3ib+1)-Pvb(pqv3ib))*(pqv3b-pqv3ib)

```

Appendix I Fortran (source file) program ia_na.f95 (Responsiveness modelling: study I)

```

! calculate iqr for v
iqr_vb = qv3b - qv1b
!calculate np srm
srmb_np(k) = vb_med / iqr_vb

!method c - drop negative cases
if (nxpos < 4) then
mc_mv(m) = mc_mv(m) + 1
go to 1 !skip over calculation of summary stats
else
xbarc = xsump/nxpos
s2ic = (xsum2p - (xsump**2)/nxpos)
s2c = s2ic/(nxpos-1)
sc = DSQRT(s2c)
srmc(k) = xbarc/sc

CALL DSORT@(vc,x,nx(m))
DO i = 1, nx(m)
  Pvc(i) = x(vc(i))
ENDDO
do i = 1, nxpos
xc(i) = Pvc(nxneg+i)
enddo

!check
!do i = 1, nx(m)
!print*, pva(i),',',xc(i)
!enddo

! print*, nx(m), nxpos, nxneg
IF (MOD(nxpos,2)==0) THEN
vc_med = (xc(nxpos/2) + xc(nxpos/2+1))/2.0
ELSE
vc_med = xc(nxpos/2+1)
END IF

! calculate position of upper and lower quartiles for v
pqv1c = ((nxpos+1)*.25)
pqv3c = ((nxpos+1)*.75)
! convert these values to integers
pqv1ic = int(pqv1c)
pqv3ic = int(pqv3c)
! calculate vb1 & vb3 (25th & 75th centile for v)
qv1c = xc(pqv1ic)+(xc(pqv1ic+1)-xc(pqv1ic))*(pqv1c-pqv1ic)
qv3c = xc(pqv3ic)+(xc(pqv3ic+1)-xc(pqv3ic))*(pqv3c-pqv3ic)
! calculate iqr for v
iqr_vc = qv3c - qv1c
!calculate np srm
srmc_np(k) = vc_med / iqr_vc

```

Appendix I Fortran (source file) program ia_na.f95 (Responsiveness modelling:
study I)

end if

! Calculate mean responsiveness statistics for simulation run
! In development raw values have been output to (e.g. spss)
! but in final modelling only the summary statistics below will
! be used.

! Calculate intermediary values for deriving parametric statistics

!based on parametic es statistics

1 esa_sum = esa_sum + srma(k)
esa_sum2 = esa_sum2 + srma(k)**2

esb_sum = esb_sum + srmb(k)
esb_sum2 = esb_sum2 + srmb(k)**2

esc_sum = esc_sum + srmc(k)
esc_sum2 = esc_sum2 + srmc(k)**2

!based on non-parametic es statistics

ena_sum = ena_sum + srma_np(k)
ena_sum2 = ena_sum2 + srma_np(k)**2

enb_sum = enb_sum + srmb_np(k)
enb_sum2 = enb_sum2 + srmb_np(k)**2

enc_sum = enc_sum + srmc_np(k)
enc_sum2 = enc_sum2 + srmc_np(k)**2

ENDDO

! calculate true denominator for method c based on
! number of valid samples (i.e. 4+ valid values per sample)
mc_val(m) = n_sim - mc_mv(m)

! calculate mean srms (p/np) for each method
esa_bar = esa_sum/n_sim
esb_bar = esb_sum/n_sim
esc_bar = esc_sum/mc_val(m)

ena_bar = ena_sum/n_sim
enb_bar = enb_sum/n_sim
enc_bar = enc_sum/mc_val(m)

! Calculate sds from p es stats
esav = (esa_sum2 - esa_sum**2/n_sim)/(n_sim-1)
esasd = DSQRT(esav)
esbv = (esb_sum2 - esb_sum**2/n_sim)/(n_sim-1)

Appendix I Fortran (source file) program ia_na.f95 (Responsiveness modelling:
study I)

```
esbsd = DSQRT(esbv)
escv = (esc_sum2 - esc_sum**2/n_sim)/(n_sim-1)
escsd = DSQRT(escv)
```

```
! Calculate sds from np es stats
enav = (ena_sum2 - ena_sum**2/n_sim)/(n_sim-1)
enasd = DSQRT(enav)
enbv = (enb_sum2 - enb_sum**2/n_sim)/(n_sim-1)
enbsd = DSQRT(enbv)
encv = (enc_sum2 - enc_sum**2/n_sim)/(n_sim-1)
encsd = DSQRT(encv)
```

! Calculate medians: parametric (each method)

```
CALL DSORT@(Aa, srma, n_sim)
DO i = 1, n_sim
  Ya(i) = srma(Aa(i))
ENDDO
IF (MOD(n_sim,2)==0) THEN
  esamd = (Ya(n_sim/2) + Ya(n_sim/2+1))/2.0
ELSE
  esamd = Ya(n_sim/2+1)
END IF
```

```
CALL DSORT@(Ab, srmb, n_sim)
DO i = 1, n_sim
  Yb(i) = srmb(Ab(i))
ENDDO
IF (MOD(n_sim,2)==0) THEN
  esbmd = (Yb(n_sim/2) + Yb(n_sim/2+1))/2.0
ELSE
  esbmd = Yb(n_sim/2+1)
END IF
```

```
CALL DSORT@(Ac, srmc, n_sim)
DO i = 1, n_sim
  Yc(i) = srmc(Ac(i))
ENDDO
IF (MOD(n_sim,2)==0) THEN
  escmd = (Yc(n_sim/2) + Yc(n_sim/2+1))/2.0
ELSE
  escmd = Yc(n_sim/2+1)
END IF
```

! Calculate medians: non-parametric

```
CALL DSORT@(Aa_np, srma_np, n_sim)
DO i = 1, n_sim
  Ya_np(i) = srma_np(Aa_np(i))
```

Appendix I Fortran (source file) program ia_na.f95 (Responsiveness modelling:
study I)

```
ENDDO
IF (MOD(n_sim,2)==0) THEN
  enamd = (Ya_np(n_sim/2) + Ya_np(n_sim/2+1))/2.0
ELSE
  enamd = Ya_np(n_sim/2+1)
END IF

CALL DSORT@(Ab_np, srmb_np, n_sim)
DO i = 1, n_sim
  Yb_np(i) = srmb_np(Ab_np(i))
ENDDO
IF (MOD(n_sim,2)==0) THEN
  enbmd = (Yb_np(n_sim/2) + Yb_np(n_sim/2+1))/2.0
ELSE
  enbmd = Yb_np(n_sim/2+1)
END IF

CALL DSORT@(Ac_np, srnc_np, n_sim)
DO i = 1, n_sim
  Yc_np(i) = srnc_np(Ac_np(i))
ENDDO
IF (MOD(n_sim,2)==0) THEN
  encmd = (Yc_np(n_sim/2) + Yc_np(n_sim/2+1))/2.0
ELSE
  encmd = Yc_np(n_sim/2+1)
END IF

! Calculate interquartile ranges
! (see also program iqr_v2)

! iqr for parametric es
espq1 = ((n_sim+1)*.25)
espq3 = ((n_sim+1)*.75)
espq1i = int(espq1)
espq3i = int(espq3)

!Calculate q1 & q3 & es iqr
esaq1=Ya(espq1i)+(Ya(espq1i+1)-Ya(espq1i))*(espq1-espq1i)
esaq3=Ya(espq3i)+(Ya(espq3i+1)-Ya(espq3i))*(espq3-espq3i)
esa_iqr = esaq3 - esaq1

esbq1=Yb(espq1i)+(Yb(espq1i+1)-Yb(espq1i))*(espq1-espq1i)
esbq3=Yb(espq3i)+(Yb(espq3i+1)-Yb(espq3i))*(espq3-espq3i)
esb_iqr = esbq3 - esbq1

escq1=Yc(espq1i)+(Yc(espq1i+1)-Yc(espq1i))*(espq1-espq1i)
escq3=Yc(espq3i)+(Yc(espq3i+1)-Yc(espq3i))*(espq3-espq3i)
esc_iqr = escq3 - escq1

! iqr for non-parametric iqr
```

Appendix I Fortran (source file) program ia_na.f95 (Responsiveness modelling:
study I)

```
enpq1 = ((n_sim+1)*.25)
enpq3 = ((n_sim+1)*.75)
enpq1i = int(enpq1)
enpq3i = int(enpq3)
```

```
!Calculate q1 / q3 / iqr
enaq1=Ya_np(enpq1i)+(Ya_np(enpq1i+1)-Ya_np(enpq1i))*(enpq1-enpq1i)
enaq3=Ya_np(enpq3i)+(Ya_np(enpq3i+1)-Ya_np(enpq3i))*(enpq3-enpq3i)
ena_iqr = enaq3 - enaq1
```

```
enbq1=Yb_np(enpq1i)+(Yb_np(enpq1i+1)-Yb_np(enpq1i))*(enpq1-enpq1i)
enbq3=Yb_np(enpq3i)+(Yb_np(enpq3i+1)-Yb_np(enpq3i))*(enpq3-enpq3i)
enb_iqr = enbq3 - enbq1
```

```
encq1=Yc_np(enpq1i)+(Yc_np(enpq1i+1)-Yc_np(enpq1i))*(enpq1-enpq1i)
encq3=Yc_np(enpq3i)+(Yc_np(enpq3i+1)-Yc_np(enpq3i))*(enpq3-enpq3i)
enc_iqr = encq3 - encq1
```

```
! Print summary responsiveness statistics
```

```
PRINT*, "
PRINT*, "
PRINT*, 'Start of simulation run using sample size:',nx(m)
print*, "
print*, 'Summary of valid / invalid simulations:'
print*, "
print*, 'valid :',mc_val(m)
print*, 'invalid:',mc_mv(m)
PRINT*, "
PRINT*, 'srm - method a',
PRINT*, ' - parametric',
PRINT*, ' (mean srm) :',esa_bar
PRINT*, ' (sd) :',esasd
PRINT*, ' (25th cntl):',esaq1
PRINT*, ' (median) :',esamd
PRINT*, ' (75th cntl):',esaq3
PRINT*, ' (iqr) :',esa_iqr
PRINT*
PRINT*, ' - non-parametric',
PRINT*, ' (mean srm) :',ena_bar
PRINT*, ' (sd) :',enasd
PRINT*, ' (25th cntl):',enaq1
PRINT*, ' (median) :',enamd
PRINT*, ' (75th cntl):',enaq3
PRINT*, ' (iqr) :',ena_iqr
PRINT*, "
PRINT*, 'srm - method b',
PRINT*, ' - parametric',
PRINT*, ' (mean srm) :',esb_bar
PRINT*, ' (sd) :',esbsd
```

Appendix I Fortran (source file) program ia_na.f95 (Responsiveness modelling:
study I)

```
PRINT*, ' (25th cntl):',esbq1
PRINT*, ' (median) :',esbmd
PRINT*, ' (75th cntl):',esbq3
PRINT*, ' (iqr) :',esb_iqr
PRINT*
PRINT*, ' - non-parametric',
PRINT*, ' (mean srm) :',enb_bar
PRINT*, ' (sd) :',enbsd
PRINT*, ' (25th cntl):',enbq1
PRINT*, ' (median) :',enbmd
PRINT*, ' (75th cntl):',enbq3
PRINT*, ' (iqr) :',enb_iqr
PRINT*, "
PRINT*, 'srm - method c',
PRINT*, ' - parametric',
PRINT*, ' (mean srm) :',esc_bar
PRINT*, ' (sd) :',escsd
PRINT*, ' (25th cntl):',escq1
PRINT*, ' (median) :',escmd
PRINT*, ' (75th cntl):',escq3
PRINT*, ' (iqr) :',esc_iqr
PRINT*
PRINT*, ' - non-parametric',
PRINT*, ' (mean srm) :',enc_bar
PRINT*, ' (sd) :',encsd
PRINT*, ' (25th cntl):',encq1
PRINT*, ' (median) :',encmd
PRINT*, ' (75th cntl):',encq3
PRINT*, ' (iqr) :',enc_iqr
PRINT*, "
```

! Calculate relative bias - parametric

```
biasa_p = (esa_bar/es)*100 - 100
```

```
biasb_p = (esb_bar/es)*100 - 100
```

```
biasc_p = (esc_bar/es)*100 - 100
```

```
PRINT*, 'bias in parametric srm - method a:', biasa_p
```

```
PRINT*, 'bias in parametric srm - method b:', biasb_p
```

```
PRINT*, 'bias in parametric srm - method c:', biasc_p
```


Appendix I Fortran (source file) program ia_na.f95 (Responsiveness modelling: study I)

! Calculate relative bias - non-parametric

```
biasa_np = (ena_bar/es_np)*100 - 100
biasb_np = (enb_bar/es_np)*100 - 100
biasc_np = (enc_bar/es_np)*100 - 100
```

```
PRINT*, 'bias in non-parametric srm - method a:', biasa_np
PRINT*, 'bias in non-parametric srm - method b:', biasb_np
PRINT*, 'bias in non-parametric srm - method c:', biasc_np
PRINT*, "
PRINT*, 'end of simulation run using sample size:',nx(m)
ENDDO
```

```
PRINT*, "
PRINT*, 'number of simulations:',n_sim
PRINT*, "
```

END ia_na

! Function to draw up values from normal distribution
! This function is called above when generating values

FUNCTION random_normal (mu,sd) RESULT (fn_val)

REAL ::fn_val

! Definition of local variables

```
REAL (KIND=2) :: RANDOM
REAL :: s = 0.449871, t = -0.386595, a = 0.19600, b = 0.25472
REAL :: r1 = 0.27597, r2 = 0.27846, half = 0.5
REAL*8 :: u, v, x, y, q, sd, mu
```

! Generate P = (u,v) uniform in rectangle enclosing acceptance region

DO

```
u = RANDOM ()
v = RANDOM ()
v = 1.7156 * (v - half)
```

! Evaluate the quadratic form

```
x = u - s
y = ABS(v) - t
q = x**2 + y*(a*y - b*x)
```

! Accept P if inside inner ellipse

```
IF (q < r1) EXIT
```

! Reject P if outside outer ellipse

```
IF (q > r2) CYCLE
```

Appendix I Fortran (source file) program ia_na.f95 (Responsiveness modelling:
study I)

```
! Reject P if outside acceptance region
  IF (v**2 < -4.0*LOG(u)*u**2) EXIT
END DO

! Return ratio of P's coordinates as the normal deviate
fn_val = (v/u)*sd+mu
RETURN

END FUNCTION random_normal
```

Appendix II Fortran (source file) program ib_nav1 (Responsiveness modelling: Study II)

```
PROGRAM ib_nav1

! Study II modelling
!
! Note: Modelling focuses upon method a (from stage 1a modelling – i.e.
! using all values as generated regardless of sign)
!
! Program based upon ia_na, and originally cmb_ds16a.ftn

! define program variables

! Note that array specifications for:
!
! x set to highest desired value and actual values are to be input
! later on in the program
!
! srm set to number of simulation runs (n=10000)

! variables for individual samples

INTEGER,dimension(6)::nx = (/15, 30, 45, 90, 180, 360/)
INTEGER nxpos, nxneg
INTEGER mc_val(6), mc_mv(6) !number of samples with lt 4 values
INTEGER:: nn = 6
REAL*8 x(360),xb(360), xc(360)! Individual change score values
REAL*8 xbara, xbarb, xbarc
REAL*8 sa, sb, sc
REAL*8,dimension(10000)::srma, srmb, srmc !sample resp. statistic
REAL*8 mux
REAL*8 sdx
REAL*8 es
REAL*8 biasa_p, biasb_p, biasc_p
REAL*8 xsuma, xsumb, xsump, xsumn
REAL*8 xsum2a, xsum2b, xsum2p, xsum2n
REAL*8 s2a, s2b, s2c
REAL*8 s2ia, s2ib, s2ic

INTEGER n_sim !number of simulations being run

! sample non-parametric statistics variables

REAL*8,dimension(10000)::pqv1a, pqv1b, pqv1c
REAL*8,dimension(10000)::pqv3a, pqv3b, pqv3c
INTEGER,dimension(10000)::pqv1ia, pqv1ib, pqv1ic
INTEGER,dimension(10000)::pqv3ia, pqv3ib, pqv3ic
REAL*8 qv1a, qv1b, qv1c
REAL*8 qv3a, qv3b, qv3c
REAL*8 iqr_va, iqr_vb, iqr_vc

INTEGER,dimension(10000)::va, vb, vc
```

Appendix II Fortran (source file) program ib_nav1 (Responsiveness modelling: Study II)

```
REAL*8,dimension(10000)::Pva, Pvb, Pvc
REAL*8 va_med, vb_med, vc_med

REAL*8,dimension(10000)::srma_np, srmb_np, srmc_np

!! simulation summary variables

! variables for calculating sd
REAL*8 esa_sum, esb_sum, esc_sum
REAL*8 ena_sum, enb_sum, enc_sum
REAL*8 esa_sum2, esb_sum2, esc_sum2
REAL*8 ena_sum2, enb_sum2, enc_sum2
REAL*8 esa_bar, esb_bar, esc_bar
REAL*8 ena_bar, enb_bar, enc_bar
REAL*8 esav, esbv, escv
REAL*8 enav, enbv, encv
REAL*8 esasd, esbsd, escsd
REAL*8 enasd, enbsd, encsd

! variables for calculating srm median
REAL*8,dimension(10000)::Ya, Yb, Yc
REAL*8,dimension(10000)::Ya_np, Yb_np, Yc_np
INTEGER,dimension(10000)::Aa, Ab, Ac
INTEGER,dimension(10000)::Aa_np, Ab_np, Ac_np
REAL*8 esamd, esbmd, escmd
REAL*8 enamd, enbmd, encmd

! variables for calculating srm interquartile range
REAL*8 espq1, espq3
REAL*8 esaq1, esaq3, esa_iqr
REAL*8 esbq1, esbq3, esb_iqr
REAL*8 escq1, escq3, esc_iqr
REAL*8 enpq1, enpq3
REAL*8 enaq1, enaq3, ena_iqr
REAL*8 enbq1, enbq3, enb_iqr
REAL*8 encq1, encq3, enc_iqr

INTEGER espq1i, espq3i
INTEGER enpq1i, enpq3i

! variables for calculating non-parametric bias estimate
REAL*8 lqrt_np, uqrt_np, iqr_np
REAL*8 es_np
REAL*8 biasa_np, biasb_np, biasc_np
REAL*8 lqrt_np1, uqrt_np1, iqr_np1

REAL*8 a1, a2, nsd, nmu, lsd, lmu
REAL*8 md_np

! set random seed
```

Appendix II Fortran (source file) program ib_nav1 (Responsiveness modelling: Study II)

```
CALL DATE_TIME_SEED@
```

```
! Section 1.3
```

```
! specify number of simulations and sample parameters
```

```
!! calculate the required mean and sd for normal distribution
```

```
lmu = 1! enter value of required log-normal mean
```

```
lsd = 10.0! enter value of required log-normal sd
```

```
a1 = 1 + (lsd / lmu)**2
```

```
a2 = log(a1)
```

```
nsd = dsqrt(a2)
```

```
nmu = log(lmu) - 0.5*a2
```

```
n_sim = 10000
```

```
mux = nmu
```

```
sdx = nsd
```

```
! calculate para/non-para effect sizes
```

```
lqrt_np1 = 0
```

```
uqrt_np1 = 0
```

```
lqrt_np = 0
```

```
uqrt_np = 0
```

```
iqr_np1 = 0
```

```
md_np = 0
```

```
es = lmu/lsd
```

```
PRINT*, 'Study I(b)'
```

```
PRINT*, "
```

```
PRINT*, 'Program series = ib_na'
```

```
PRINT*, "
```

```
PRINT*, 'Date: 8th Feb 2005'
```

```
PRINT*, "
```

```
PRINT*, 'Process statistics'
```

```
PRINT*, "
```

```
PRINT*, 'specified log normal parameters:'
```

```
PRINT*, ' - mean change      :',lmu
```

```
PRINT*, ' - standard deviation:',lsd
```

```
PRINT*, "
```

```
lqrt_np1 = nmu - (nsd*0.67449)
```

```
uqrt_np1 = nmu + (nsd*0.67449)
```

```
PRINT*, 'lq for normal distribution:', lqrt_np1
```

```
PRINT*, 'uq for normal distribution:', uqrt_np1
```

```
lqrt_np = exp(lqrt_np1)
```

```
uqrt_np = exp(uqrt_np1)
```

```
md_np = exp(nmu)
```

```
PRINT*, 'lq for log-transformed distribution:', lqrt_np
```

Appendix II Fortran (source file) program ib_nav1 (Responsiveness modelling: Study II)

```
PRINT*, 'uq for log-transformed distribution:', uqrt_np
PRINT*, "
iqr_np1 = uqrt_np1 - lqrt_np1
es_np1 = mux / iqr_np1
!PRINT*, 'iqr_np1:', iqr_np1
!PRINT*, 'es_np1 :', es_np1

iqr_np = uqrt_np - lqrt_np
es_np = md_np / iqr_np
PRINT*, 'md_np :', md_np
PRINT*, 'iqr_np:', iqr_np
PRINT*, "
PRINT*, 'This simulation uses the following true SRMs:'
PRINT*, "
PRINT*, ' - pSRM :', es
PRINT*, ' - npSRM:', es_np

! Notes about non-parametric effect sizes
!
! For assessment of bias, the derived es is compared with the 'true' value. The
! 'true value' is calculated from given parameters - the mean and sd of the
! distribution from which values are generated.

! For parametric es, these true values are simply those specified by the
modelling
! plan and result in ess which range from 0.1 to 2.

! For non-parametric ess, the 'true es' is calculated from the mean change
! (again given), and the interquartile range. The latter has to be calculated, and
! involves firstly calculating the values of the upper and lower quartiles.
!
! For data generated from a normal distribution, the quartiles are calculated
! by applying a factor to the population sd (and adding/subtracting from the
! population mean).
!
! For log-normal data, the quartiles are calculated

! Start simulation

Do m = 1, nn

esa_sum = 0
esb_sum = 0
esc_sum = 0
ena_sum = 0
enb_sum = 0
enc_sum = 0
esa_sum2 = 0
esb_sum2 = 0
esc_sum2 = 0
```

Appendix II Fortran (source file) program ib_nav1 (Responsiveness modelling: Study II)

```
ena_sum2 = 0
enb_sum2 = 0
enc_sum2 = 0
mc_mv(m) = 0
```

```
DO k = 1, n_sim
```

```
xsuma = 0
xsumb = 0
xsump = 0
xsumn = 0
xsum2a = 0
xsum2b = 0
xsum2p = 0
xsum2n = 0
xbara = 0
xbarb = 0
s2ia = 0
s2ib = 0
s2ic = 0
s2a = 0
s2b = 0
s2c = 0
sa = 0
sb = 0
sc = 0
nxpos = 0
nxneg = 0
va_med = 0
vb_med = 0
vc_med = 0
```

```
! Generate change values
```

```
DO i = 1, nx(m)
  x(i) = random_normal(mux, sdx)
  x(i) = exp(x(i))
  xb(i) = x(i)
  if (xb(i) < 0) then
    xb(i) = (xb(i)*(-1))
  end if

  if (x(i)<0) then
    xsumn = xsumn + x(i)
    xsum2n = xsum2n + x(i)**2
    nxneg = nxneg + 1
  else
    xsump = xsump + x(i)
    xsum2p = xsum2p + x(i)**2
  endif
ENDDO
```

Appendix II Fortran (source file) program ib_nav1 (Responsiveness modelling: Study II)

```
nxpos = nx(m)-nxneg
```

```
!calculate sample mean, sd and median for each of the three methods
```

```
!method a - all cases, as generated
```

```
xsuma = xsump + xsumn
```

```
xsum2a = xsum2p + xsum2n
```

```
xbara = xsuma /nx(m)
```

```
s2ia = (xsum2a - (xsuma**2)/nx(m))
```

```
s2a = s2ia/(nx(m)-1)
```

```
sa = DSQRT(s2a)
```

```
srma(k) = xbara /sa
```

```
CALL DSORT@(va,x,nx(m))
```

```
DO i = 1, nx(m)
```

```
    Pva(i) = x(va(i))
```

```
ENDDO
```

```
IF (MOD(nx(m),2)==0) THEN
```

```
    va_med = (Pva(nx(m)/2) + Pva(nx(m)/2+1))/2.0
```

```
ELSE
```

```
    va_med = Pva(nx(m)/2+1)
```

```
END IF
```

```
! calculate position of upper and lower quartiles for v
```

```
pqv1a = ((nx(m)+1)*.25)
```

```
pqv3a = ((nx(m)+1)*.75)
```

```
! convert these values to integers
```

```
pqv1ia = int(pqv1a)
```

```
pqv3ia = int(pqv3a)
```

```
! calculate vb1 & vb3 (25th & 75th centile for v)
```

```
qv1a = Pva(pqv1ia)+(Pva(pqv1ia+1)-Pva(pqv1ia))*(pqv1a-pqv1ia)
```

```
qv3a = Pva(pqv3ia)+(Pva(pqv3ia+1)-Pva(pqv3ia))*(pqv3a-pqv3ia)
```

```
! calculate iqr for v
```

```
iqr_va = qv3a - qv1a
```

```
!calculate np srm
```

```
srma_np(k) = va_med / iqr_va
```

```
!method b - all cases, negative sign reversed
```

```
xsumb = xsump + ((xsumn)*(-1))
```

```
xsum2b = xsum2n + xsum2p
```

```
xbarb = xsumb/nx(m)
```

```
s2ib = (xsum2b - (xsumb**2)/nx(m))
```

```
s2b = s2ib/(nx(m)-1)
```

```
sb = DSQRT(s2b)
```

```
srmb(k) = xbarb/sb
```

```
CALL DSORT@(vb,xb,nx(m))
```

```
DO i = 1, nx(m)
```

```
    Pvb(i) = xb(vb(i))
```


Appendix II Fortran (source file) program ib_nav1 (Responsiveness modelling: Study II)

```

!Print*, pvb(i)
  ENDDO
IF (MOD(nx(m),2)==0) THEN
  vb_med = (Pvb(nx(m)/2) + Pvb(nx(m)/2+1))/2.0
ELSE
  vb_med = Pvb(nx(m)/2+1)
END IF

! calculate position of upper and lower quartiles for v
pqv1b = ((nx(m)+1)*.25)
pqv3b = ((nx(m)+1)*.75)
! convert these values to integers
pqv1ib = int(pqv1b)
pqv3ib = int(pqv3b)
! calculate vb1 & vb3 (25th & 75th centile for v)
qv1b = Pvb(pqv1ib)+(Pvb(pqv1ib+1)-Pvb(pqv1ib))*(pqv1b-pqv1ib)
qv3b = Pvb(pqv3ib)+(Pvb(pqv3ib+1)-Pvb(pqv3ib))*(pqv3b-pqv3ib)
! calculate iqr for v
iqr_vb = qv3b - qv1b
!calculate np srm
srmb_np(k) = vb_med / iqr_vb

!method c - drop negative cases
if (nxpos < 4) then
  mc_mv(m) = mc_mv(m) + 1
  go to 1 !skip over calculation of summary stats
else
  xbarc = xsump/nxpos
  s2ic = (xsum2p - (xsump**2)/nxpos)
  s2c = s2ic/(nxpos-1)
  sc = DSQRT(s2c)
  srmc(k) = xbarc/sc

CALL DSORT@(vc,x,nx(m))
DO i = 1, nx(m)
  Pvc(i) = x(vc(i))
ENDDO
do i = 1, nxpos
xc(i) = Pvc(nxneg+i)
enddo

IF (MOD(nxpos,2)==0) THEN
  vc_med = (xc(nxpos/2) + xc(nxpos/2+1))/2.0
ELSE
  vc_med = xc(nxpos/2+1)
END IF

! calculate position of upper and lower quartiles for v
pqv1c = ((nxpos+1)*.25)
pqv3c = ((nxpos+1)*.75)

```

Appendix II Fortran (source file) program ib_nav1 (Responsiveness modelling: Study II)

```
! convert these values to integers
pqv1c = int(pqv1c)
pqv3c = int(pqv3c)
! calculate vb1 & vb3 (25th & 75th centile for v)
qv1c = xc(pqv1c)+(xc(pqv1c+1)-xc(pqv1c))*(pqv1c-pqv1c)
qv3c = xc(pqv3c)+(xc(pqv3c+1)-xc(pqv3c))*(pqv3c-pqv3c)
! calculate iqr for v
iqr_vc = qv3c - qv1c
!calculate np srm
srnc_np(k) = vc_med / iqr_vc

end if

! Calculate mean responsiveness statistics for simulation run
! In development raw values have been output to (e.g. spss)
! but in final modelling only the summary statistics below will
! be used.

! Calculate intermediary values for deriving parametric statistics

!based on parametic es statistics

1  esa_sum = esa_sum + srma(k)
   esa_sum2 = esa_sum2 + srma(k)**2

   esb_sum = esb_sum + srmb(k)
   esb_sum2 = esb_sum2 + srmb(k)**2

   esc_sum = esc_sum + srnc(k)
   esc_sum2 = esc_sum2 + srnc(k)**2

!based on non-parametic es statistics

ena_sum = ena_sum + srma_np(k)
ena_sum2 = ena_sum2 + srma_np(k)**2

enb_sum = enb_sum + srmb_np(k)
enb_sum2 = enb_sum2 + srmb_np(k)**2

enc_sum = enc_sum + srnc_np(k)
enc_sum2 = enc_sum2 + srnc_np(k)**2

!check srm vales (parametric)

ENDDO

!check
!Do i = 1,n_sim
! PRINT*, 'run:',i
! PRINT*, 'mean  :', xbar(i)
```

Appendix II Fortran (source file) program ib_nav1 (Responsiveness modelling: Study II)

```

! PRINT*, 'xsum   :', xsum(i)
! PRINT*, 'xsum2  :', xsum2(i)
! PRINT*, 's2i    :', s2i(i)
! PRINT*, 's2     :', s2(i)
! PRINT*, 'sd     :', s(i)
! PRINT*, 'srm (p) :', srm(i)
! PRINT*, 'lwr qrtile:', qv1(i)
! PRINT*, 'median  :', v_med(i)
! PRINT*, 'upr qrtile:', qv3(i)
! PRINT*, 'iqr     :', iqr_v(i)
! PRINT*, 'srm (np) :', srm_np(i)
!enddo

! calculate true denominator for method c based on
! number of valid samples (i.e. 4+ valid values per sample)
mc_val(m) = n_sim - mc_mv(m)

! calculate mean srms (p/np) for each method
esa_bar = esa_sum/n_sim
esb_bar = esb_sum/n_sim
esc_bar = esc_sum/mc_val(m)

ena_bar = ena_sum/n_sim
enb_bar = enb_sum/n_sim
enc_bar = enc_sum/mc_val(m)

! Calculate sds from p es stats
esav = (esa_sum2 - esa_sum**2/n_sim)/(n_sim-1)
esasd = DSQRT(esav)
esbv = (esb_sum2 - esb_sum**2/n_sim)/(n_sim-1)
esbsd = DSQRT(esbv)
escv = (esc_sum2 - esc_sum**2/n_sim)/(n_sim-1)
escsd = DSQRT(escv)

! Calculate sds from np es stats
enav = (ena_sum2 - ena_sum**2/n_sim)/(n_sim-1)
enasd = DSQRT(enav)
enbv = (enb_sum2 - enb_sum**2/n_sim)/(n_sim-1)
enbsd = DSQRT(enbv)
encv = (enc_sum2 - enc_sum**2/n_sim)/(n_sim-1)
encsd = DSQRT(encv)

! Calculate medians: parametric (each method)

CALL DSORT@(Aa, srma, n_sim)
DO i = 1, n_sim
  Ya(i) = srma(Aa(i))
ENDDO
IF (MOD(n_sim,2)==0) THEN
  esamd = (Ya(n_sim/2) + Ya(n_sim/2+1))/2.0

```

Appendix II Fortran (source file) program ib_nav1 (Responsiveness modelling: Study II)

```
ELSE
  esamd = Ya(n_sim/2+1)
END IF

CALL DSORT@(Ab, srmb, n_sim)
DO i = 1, n_sim
  Yb(i) = srmb(Ab(i))
ENDDO
IF (MOD(n_sim,2)==0) THEN
  esbmd = (Yb(n_sim/2) + Yb(n_sim/2+1))/2.0
ELSE
  esbmd = Yb(n_sim/2+1)
END IF

CALL DSORT@(Ac, srmc, n_sim)
DO i = 1, n_sim
  Yc(i) = srmc(Ac(i))
ENDDO
IF (MOD(n_sim,2)==0) THEN
  escmd = (Yc(n_sim/2) + Yc(n_sim/2+1))/2.0
ELSE
  escmd = Yc(n_sim/2+1)
END IF

! Calculate medians: non-parametric

CALL DSORT@(Aa_np, srma_np, n_sim)
DO i = 1, n_sim
  Ya_np(i) = srma_np(Aa_np(i))
ENDDO
IF (MOD(n_sim,2)==0) THEN
  enamd = (Ya_np(n_sim/2) + Ya_np(n_sim/2+1))/2.0
ELSE
  enamd = Ya_np(n_sim/2+1)
END IF

CALL DSORT@(Ab_np, srmb_np, n_sim)
DO i = 1, n_sim
  Yb_np(i) = srmb_np(Ab_np(i))
ENDDO
IF (MOD(n_sim,2)==0) THEN
  enbmd = (Yb_np(n_sim/2) + Yb_np(n_sim/2+1))/2.0
ELSE
  enbmd = Yb_np(n_sim/2+1)
END IF

CALL DSORT@(Ac_np, srmc_np, n_sim)
DO i = 1, n_sim
  Yc_np(i) = srmc_np(Ac_np(i))
ENDDO
```

Appendix II Fortran (source file) program ib_nav1 (Responsiveness modelling: Study II)

```
IF (MOD(n_sim,2)==0) THEN
  encmd = (Yc_np(n_sim/2) + Yc_np(n_sim/2+1))/2.0
ELSE
  encmd = Yc_np(n_sim/2+1)
END IF

! Calculate interquartile ranges
! (see also program iqr_v2)

! iqr for parametric es
espq1 = ((n_sim+1)*.25)
espq3 = ((n_sim+1)*.75)
espq1i = int(espq1)
espq3i = int(espq3)

!Calculate q1 & q3 & es iqr
esaq1=Ya(espq1i)+(Ya(espq1i+1)-Ya(espq1i))*(espq1-espq1i)
esaq3=Ya(espq3i)+(Ya(espq3i+1)-Ya(espq3i))*(espq3-espq3i)
esa_iqr = esaq3 - esaq1

esbq1=Yb(espq1i)+(Yb(espq1i+1)-Yb(espq1i))*(espq1-espq1i)
esbq3=Yb(espq3i)+(Yb(espq3i+1)-Yb(espq3i))*(espq3-espq3i)
esb_iqr = esbq3 - esbq1

escq1=Yc(espq1i)+(Yc(espq1i+1)-Yc(espq1i))*(espq1-espq1i)
escq3=Yc(espq3i)+(Yc(espq3i+1)-Yc(espq3i))*(espq3-espq3i)
esc_iqr = escq3 - escq1

! iqr for non-parametric iqr
enpq1 = ((n_sim+1)*.25)
enpq3 = ((n_sim+1)*.75)
enpq1i = int(enpq1)
enpq3i = int(enpq3)

!Calculate q1 / q3 / iqr
enaq1=Ya_np(enpq1i)+(Ya_np(enpq1i+1)-Ya_np(enpq1i))*(enpq1-enpq1i)
enaq3=Ya_np(enpq3i)+(Ya_np(enpq3i+1)-Ya_np(enpq3i))*(enpq3-enpq3i)
ena_iqr = enaq3 - enaq1

enbq1=Yb_np(enpq1i)+(Yb_np(enpq1i+1)-Yb_np(enpq1i))*(enpq1-enpq1i)
enbq3=Yb_np(enpq3i)+(Yb_np(enpq3i+1)-Yb_np(enpq3i))*(enpq3-enpq3i)
enb_iqr = enbq3 - enbq1

encq1=Yc_np(enpq1i)+(Yc_np(enpq1i+1)-Yc_np(enpq1i))*(enpq1-enpq1i)
encq3=Yc_np(enpq3i)+(Yc_np(enpq3i+1)-Yc_np(enpq3i))*(enpq3-enpq3i)
enc_iqr = encq3 - encq1
```

Appendix II Fortran (source file) program ib_nav1 (Responsiveness modelling: Study II)

! Print summary responsiveness statistics

```
PRINT*, "  
PRINT*, "  
PRINT*, 'Start of simulation run using sample size:',nx(m)  
print*, "  
print*, 'Summary of valid / invalid simulations:'  
print*, "  
print*, 'valid :',mc_val(m)  
print*, 'invalid:',mc_mv(m)  
PRINT*, "  
PRINT*, 'srm',  
PRINT*, ' - parametric',  
PRINT*, ' (mean srm) :',esa_bar  
PRINT*, ' (sd) :',esasd  
PRINT*, ' (25th cntl):',esaq1  
PRINT*, ' (median) :',esamd  
PRINT*, ' (75th cntl):',esaq3  
PRINT*, ' (iqr) :',esa_iqr  
PRINT*  
PRINT*, ' - non-parametric',  
PRINT*, ' (mean srm) :',ena_bar  
PRINT*, ' (sd) :',enasd  
PRINT*, ' (25th cntl):',enaq1  
PRINT*, ' (median) :',enamd  
PRINT*, ' (75th cntl):',enaq3  
PRINT*, ' (iqr) :',ena_iqr  
PRINT*, "  
!PRINT*, 'srm - method b',  
!PRINT*, ' - parametric',  
!PRINT*, ' (mean srm) :',esb_bar  
!PRINT*, ' (sd) :',esbsd  
!PRINT*, ' (25th cntl):',esbq1  
!PRINT*, ' (median) :',esbmd  
!PRINT*, ' (75th cntl):',esbq3  
!PRINT*, ' (iqr) :',esb_iqr  
!PRINT*  
!PRINT*, ' - non-parametric',  
!PRINT*, ' (mean srm) :',enb_bar  
!PRINT*, ' (sd) :',enbsd  
!PRINT*, ' (25th cntl):',enbq1  
!PRINT*, ' (median) :',enbmd  
!PRINT*, ' (75th cntl):',enbq3  
!PRINT*, ' (iqr) :',enb_iqr  
!PRINT*, "  
!PRINT*, 'srm - method c',  
!PRINT*, ' - parametric',  
!PRINT*, ' (mean srm) :',esc_bar  
!PRINT*, ' (sd) :',escsd  
!PRINT*, ' (25th cntl):',escq1
```

Appendix II Fortran (source file) program ib_nav1 (Responsiveness modelling: Study II)

```

!PRINT*, ' (median) :',escmd
!PRINT*, ' (75th cntl):',escq3
!PRINT*, ' (iqr) :',esc_iqr
!PRINT*
!PRINT*, ' - non-parametric',
!PRINT*, ' (mean srm) :',enc_bar
!PRINT*, ' (sd) :',encsd
!PRINT*, ' (25th cntl):',encq1
!PRINT*, ' (median) :',encmd
!PRINT*, ' (75th cntl):',encq3
!PRINT*, ' (iqr) :',enc_iqr
!PRINT*, "

! Calculate relative bias - parametric

biasa_p = (esa_bar/es)*100 - 100
biasb_p = (esb_bar/es)*100 - 100
biasc_p = (esc_bar/es)*100 - 100

PRINT*, 'bias in parametric srm - method a:', biasa_p
!PRINT*, 'bias in parametric srm - method b:', biasb_p
!PRINT*, 'bias in parametric srm - method c:', biasc_p

! Calculate relative bias - non-parametric

biasa_np = (ena_bar/es_np)*100 - 100
biasb_np = (enb_bar/es_np)*100 - 100
biasc_np = (enc_bar/es_np)*100 - 100

PRINT*, 'bias in non-parametric srm - method a:', biasa_np
!PRINT*, 'bias in non-parametric srm - method b:', biasb_np
!PRINT*, 'bias in non-parametric srm - method c:', biasc_np
PRINT*, "
PRINT*, 'end of simulation run using sample size:',nx(m)
ENDDO

PRINT*, "
PRINT*, 'number of simulations:',n_sim
PRINT*, "

END ib_nav1

! Function to draw up values from normal distribution
! This function is called above when generating values

FUNCTION random_normal (mu,sd) RESULT (fn_val)

REAL ::fn_val

! Definition of local variables

```

Appendix II Fortran (source file) program ib_nav1 (Responsiveness modelling: Study II)

```
REAL (KIND=2) :: RANDOM
REAL :: s = 0.449871, t = -0.386595, a = 0.19600, b = 0.25472
REAL :: r1 = 0.27597, r2 = 0.27846, half = 0.5
REAL*8 :: u, v, x, y, q, sd, mu

! Generate P = (u,v) uniform in rectangle enclosing acceptance region

DO
  u = RANDOM ()
  v = RANDOM ()
  v = 1.7156 * (v - half)

! Evaluate the quadratic form
  x = u - s
  y = ABS(v) - t
  q = x**2 + y*(a*y - b*x)

! Accept P if inside inner ellipse
  IF (q < r1) EXIT
! Reject P if outside outer ellipse
  IF (q > r2) CYCLE
! Reject P if outside acceptance region
  IF (v**2 < -4.0*LOG(u)*u**2) EXIT
END DO

! Return ratio of P's coordinates as the normal deviate
fn_val = (v/u)*sd+mu
RETURN

END FUNCTION random_normal
```


Appendix III Background publications from response shift literature review (featuring review, discussion or methodological consideration of response shift)

First author	Year	Outcome addressed	Context	Patient / clinical group	Comments	Paper
Eiser	2004	Quality of Life	Clinical	Paediatrics	Reviews the place of quality of life assessment in clinical trials in paediatrics. Considers the properties of cancer-specific scales and describes some of the barriers to the use of quality of life measures in trials. Amongst the latter includes cost, maturation and response shift. Describes social comparison as one mechanism by which response shift may operate.	159
Lowy	2004	Quality of Life	Clinical	Cancer	Introduces a multilevel modelling approach to the assessment of re-conceptualisation response shift (which the authors consider incorporates both re-prioritisation and re-conceptualisation). Method tested with empirical data and in modelling studies.	157
Rapkin	2004	Quality of Life	Clinical	Various	Presents revised and testable model of response shift with focus upon assessing the quality of life appraisal process.	95
Schwartz	2004	Quality of Life	Clinical	Various	Presents a reconsideration of the meaning, value and use of psychometrics in quality of life research in the light of theoretical development of the response shift model and accumulated empirical work. Companion paper to Rapkin and Schwartz (2004). ⁹⁵	65
Norman	2003	Quality of Life	Clinical	Non-specific	Compares two theoretical perspectives on the retrospective assessment of change – response shift and implicit theories – and the validity of assumptions made in each case.	110
Ubel	2003	Quality of Life / health utility	Clinical	Various	Reviews various causes of discrepancy between public estimates of health utility and patient reported utility. Response shift (specifically scale re-calibration) is described within this as an example of assessments being made using different 'measuring sticks'.	158
Brossart	2002	Quality of Life	Clinical	Paediatrics	Reviews conceptual models of response shift and related constructs, with a particular interest in the potential impact upon the interpretation of longitudinal studies. Reference in particular is made to paediatrics and factors that may be relevant to that population. Introduces growth modelling as a statistical method for detecting the timing of occurrence of response shift and assessing different forms of response shift (shape). Considers other factors which may affect the internal validity of longitudinal studies (which should be assessed prior to modelling).	111
Carr	2002	Quality of Life, pain, function	Clinical	Rheumatoid / osteoarthritis	Reviews briefly the role of response shift in confounding assessment in trials especially within the setting of osteo- and rheumatoid arthritis.	317

First author	Year	Outcome addressed	Context	Patient / clinical group	Comments	Paper
Mick	2002	Quality of Life	Clinical	Elderly	Reviews theories of aging and adaptation and includes a consideration of the role of response shift in the clinical and social management of elderly individuals using a case study.	318
Robling	2002	Pain	Clinical	Chronic pain	Reviews different explanations for observed results in empirical study including response shift.	57
Schwartz	2002	Quality of Life	Clinical	Cancer	Reviews developments in the conception, definition, assessment and application of quality of life in clinical practice and research. Specifically it assesses the contribution of quality of life study to oncology. It highlights response shift as one of two areas for future development in clinical quality assurance and research. It also notes some of the methodological problems inherent within some design based studies.	160
Sprangers	2002	HRQoL	Clinical	Cancer	Reviews the role, achievements and methodological challenges of HRQoL assessment in oncology. Under the latter Sprangers assesses the importance of responses shift and describes some of the methodologies being used to assess it, most notably the retrospective pretest-posttest method (thentest).	237
Adang	2001	HRQoL, cost-effectiveness	Clinical	Diabetes, ESRD	Provides further comment and explanation of the authors view following a previous study of response shift in patients undergoing pancreas-kidney transplant. They had previously argued that prospective assessment of (conventional) change was the preferred study design and not a retrospectively derived change score (using the thentest) for cost-effectiveness assessment. In this paper they leave open for debate whether assessment should be made in the certainty of knowledge of actual outcome (thentest) or baseline uncertainty of outcome (conventional pre-test).	171
Carver	2000	Quality of Life	Clinical	General	Describes self-regulatory feedback systems for goal-seeking behaviour and affect management which are required for normal adaptive self-regulation. They consider that these normal mechanisms may account for the phenomena that are described as response shift.	108

Appendix III Background publications from response shift literature review (featuring review, discussion or methodological consideration of response shift)

First author	Year	Outcome addressed	Context	Patient / clinical group	Comments	Paper
Llewellyn-Thomas	2000	Quality of Life	Clinical	General	Book chapter describing sources of artefact in study designs that may confound the identification of each form of response shift (re-calibration, re-prioritisation, re-conceptualisation) when patients are evaluating health states. Examples of sources of artefact include item framing effects, sequencing effects and format effects.	265
Parducci	2000	General	General	Not applicable	Describes a broad historical background to response shift as an introduction to the Schwartz and Sprangers book 'Adaptation to Changing Health'.	319
Richards	2000	Psychological well-being	Clinical	Caregivers	Describes response shift as a result of specific meaning-based coping processes that operate when previous beliefs, expectations and goals are no longer tenable. They provide a conceptualisation of coping behaviour that is consistent with the Sprangers and Schwartz model of coping as one mechanism supporting response shift. They illustrate their discussion with data drawn from a study of caregivers to patients with AIDS who were interviewed over a five year period.	109
Schwartz	2000	Quality of Life	Clinical	General	Discussion of various theoretical papers on response shift and some of the conceptual issues that had been highlighted. These included the inter-relationship between elements of response shift (and, for example, whether response shift is an active or passive process); certain circularities / similarities in components in the response shift model (e.g. similar operationalisation of mechanisms and response shift); certain definitional issues (and, for example, whether response shift may occur simply with the passage of time).	96
Schwartz	2000	Quality of life	Clinical	Various	Book chapter summarising the implications of response shift within clinical research, drawing together various examples from other chapters in the Schwartz and Sprangers book. For example, discusses how studying response shift may illuminate the nature of quality of life experience over time; the use of response shift as an intentional clinical intervention; consideration of response shift when assessing trial results; and the role of response shift in medical decision-making.	189

First author	Year	Outcome addressed	Context	Patient / clinical group	Comments	Paper
Sprangers	2000	Quality of life	Clinical	General	Discussion section reviewing preceding chapters reporting methodological approaches to response shift assessment. In particular, points to some key methodological challenges (e.g. inter-relatedness between assessment method and response shift; identification of suitable control groups; method triangulation; and the role of secondary data analysis).	262
Gibbons	1999	Various	Clinical / various	Various	Review social and cognitive psychological literature on social comparison and the role that this may play in mediating response shift in response to changes in health status.	101
Schwartz	1999	Quality of life	Clinical	General	Introduction to the symposium on response shift reprinted in the special issue of <i>Social Science & Medicine</i> .	114
Schwartz	1999	Quality of life	Clinical	General	Reviews and recommends methodological approaches for the study of response shift in longitudinal health-related quality of life research. Delineates individualised; preference-based; successive comparison; design; statistical; and qualitative approaches. Each approach is assessed according to criteria of feasibility, reliability, validity and whether empirical data was currently available. Also discusses the interconnected nature of response shift components and the challenges posed for assessment.	11
Sprangers	1999	Quality of life	Clinical	General	Presents the theoretical model of response shift as it may affect health-related quality of life as a result of changes in health. Incorporates the components: a catalyst; antecedents; mechanisms; response shift and perceived quality of life.	10
Sprangers	1999	Quality of Life	Clinical	Cancer	Editorial introducing a paper about reframing of perception amongst cancer patients. Briefly describes the importance of response shift, reviews some major clinical findings in the field and discusses some methodological implications, including methods for assessing response shift.	166
Wilson	1999	Quality of life	Clinical	General	Discussion paper reviewing the role of response shift in the clinical setting. It proposes how response shift may address certain common clinical problems such as somatisation, hypochondria and placebo effects. It further discusses the role that response shift may play as a valid clinical intervention when biological and physiological change has been exhausted. Calls for further descriptive work and emphasises social support and coping mechanisms.	161

Appendix III Background publications from response shift literature review (featuring review, discussion or methodological consideration of response shift)

First author	Year	Outcome addressed	Context	Patient / clinical group	Comments	Paper
Smets	1993	Fatigue	Clinical	Cancer	Review of the definition, measurement and mechanisms of fatigue as experienced by patients with cancer. Introduces response shift as a potential explanatory mechanism for explaining similarities in reported levels of fatigue in patient and general population studies, and in cancer patients in before and after studies.	170
Breetvelt	1991	Quality of Life (emotional distress)	Clinical	Cancer	Describes the 'underreporting' of emotional distress by cancer patients using empirical and theoretical examples, reviews theories account for coping / adaptation to physical illness including an introduction to response shift.	320
Koele	1988	Not specified	Not specified	Not specified	Presents a method for analysing retrospective pre-test /post-test designs with a control group, incorporating a series of hierarchical models which examine and adjust for the presence of treatment and response bias effect.	321
Sprangers	1988	Various	Education / training	Various	Literature review of experimental studies employing a thentest design and citing original paper of Howard et al, 1979. Aim of the review was to evaluate subject bias as an alternative explanation to response shift in reported studies. Whilst in general this alternative explanation was not supported, there was still some evidence that subject bias may play some role in thentest designs	322

Notes re: background publications

The following sources were searched to identify the above papers: Ovid Medline (1966 to February Week 2, 2004), Embase (1980 to 2004 Week 08); PsycInfo (1985 to Feb Week 3 2004) using the phrase response shift.

Additionally two text books were identified but not included in the above review:

- Sprangers, M. (1998) Response shift and the retrospective pretest: On the usefulness of retrospective pretest-posttest designs in detecting training related response shifts. Foundation for Educational Research, Amsterdam
- Schwartz, Carolyn E. Design issues for clinical research in health psychology. Chapter in: Johnston DW, Johnston M (Eds). (2001) Health Psychology, Volume 8. Comprehensive clinical psychology. Pergamon Press: Oxford.

Appendix IV Patient approach letter (Radiology)

Date as postmark

Dear

We are writing to you because of your referral to the Department of Radiology for your continuing knee problem. The Department of Radiology is taking part in a large research project to find out the best way of treating knee problems. The research will help doctors treat patients with knee problems in the future.

We are particularly interested in the effect your knee problem may have on your own well-being. Therefore, we would like to ask if our researcher could visit to talk with you about your knee problem. This will give you an opportunity to tell us how you are affected by your condition. This should help us to understand what issues are important for people who are experiencing difficulties with their knee. We may ask you to see you again after a few months to talk with you about your knee problem and how it is affecting you.

We hope that you will feel able to take part in the research. If you do take part we will tell your doctor so that they are aware of this. We have enclosed an information sheet that describes the study in more detail. We hope that it will answer some of the questions you may have about taking part. However, if you would like to ask any other questions about the research, please feel free to get in touch with the researcher, Mike Robling (Tel: 029 2073 3018), who will be happy to answer your questions.

If you do want to take part can you please return the accompanying form. While your help in this project would be greatly appreciated, it is completely voluntary. If you do not want to take part, it will not affect the care you receive in any way.

Many thanks for your help.

Yours sincerely

Dr Kathleen Lyons

Consultant Radiologist

Direct Access to Magnetic resonance imaging: Assessment for Suspect Knees (DAMASK)

A randomised trial in England, Wales and Scotland, funded by the Medical Research Council

(International Standardised Randomised Controlled Trial Number 52135255)

Date as postmark

Dear

We are writing to you because of your referral to the Department of Orthopaedics for your continuing knee problem. The Department of Orthopaedics is taking part in a large research project to find out the best way of treating knee problems. The research will help doctors treat patients with knee problems in the future.

We are particularly interested in the effect your knee problem may have on your own well-being. Therefore, we would like to ask if our researcher could visit to talk with you about your knee problem. This will give you an opportunity to tell us how you are affected by your condition. This should help us to understand what issues are important for people who are experiencing difficulties with their knee. We may ask you to see you again after a few months to talk with you about your knee problem and how it is affecting you.

We hope that you will feel able to take part in the research. If you do take part we will tell your doctor so that they are aware of this. We have enclosed an information sheet that describes the study in more detail. We hope that it will answer some of the questions you may have about taking part. However, if you would like to ask any other questions about the research, please feel free to get in touch with the researcher, Mike Robling (Tel: 029 2073 3018), who will be happy to answer your questions.

If you do want to take part can you please return the accompanying form. While your help in this project would be greatly appreciated, it is completely voluntary. If you do not want to take part, it will not affect the care you receive in any way.

Many thanks for your help.

Yours sincerely

Mr Chris Wilson

Consultant Orthopaedic Surgeon

Study of referral to magnetic resonance imaging or orthopaedics for GP patients with knee problems, funded by the Medical Research Council

Invitation to participate

You are being invited to take part in a research study funded by the Medical Research Council (MRC) and Department of Health (DH). Before you decide, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for reading this.

Why do you want me to take part in this study?

You are one of many people who see their family doctor about problems with their knee. Your doctor can refer you to a consultant orthopaedic surgeon or to have an investigation called magnetic resonance imaging (MRI) which has been shown to be useful in identifying problems in the knee for some patients. The result of the MRI scan can help them decide whether to refer a patient to the surgeon. If they do refer, the MRI result may help the surgeon decide how quickly the patient should be seen and what treatment the patient needs. If the MRI result shows no serious problem, the family doctor could treat the patient without them having to wait to see a consultant. However, we are uncertain about the best way to use MRI. This study is therefore important to establish how best to use MRI for patients with continuing knee problems.

To get full value from our study we need to know how patients knee problems change over time. Therefore we need to develop a questionnaire for assessing the symptoms of knee problems and its treatment.

Why have I been chosen?

The study will involve patients with knee problems who have been referred by their family doctor to have an MRI scan or be seen by an orthopaedic surgeon.

What will I do if I take part?

If you agree to take part you will be asked to have a one-to-one discussion with a researcher to identify health issues that are important to you in the context of your knee problem and its management. We may also ask to see you again after a few months to talk with you about your knee problem and how it is affecting you.

The study team will not tell anyone else about what is discussed. Nor shall we tell anyone else what you tell your doctor.

What are the benefits of taking part?

The main benefit is that care of patients with knee problems should be better in the future, both for other patients and for you if you get the same problem again.

Do I have to take part?

No - the choice is yours. If you do not want to take part you do not have to give a reason. Your decision will not influence the care you receive in any way. If you choose to take part, you can still withdraw at any time if you change your mind.

What do I do now?

Thank you for thinking about taking part in this study. If you would like to take part please complete the consent form and return it to us using the self-addressed envelope.

If you would like to ask any other questions about the research please feel free to get in touch with Mike Robling (Tel: 029 2073 3018), who will be happy to answer your questions.

CONSENT FORM

Study Number:

Participant Identification Number:

Title of Project: Study of referral to magnetic resonance imaging or orthopaedics for GP patients with knee problems

Names of researcher: Mike Robling (Tel: 029 2073 3018)

I confirm that I have read and understand the information sheet for the above study and have had the opportunity to ask questions.

I understand that my participation is voluntary and that I am free to withdraw at any time.

I know that I may ask now, or in the future, any questions that I have about the study or the research procedures.

I know that records relating to me will be kept confidential. No information will be released or printed that would identify me without my permission unless required by law.

I hereby consent to participate in this study

Signature

Name (Print)

Date

Contact telephone number:

Address:

Appendix XIII Follow-up interview schedule (incorporating SEIQoL-DW and EQ-5D)

Follow-up interview schedule ID no: Date:

ii) Determining levels

Now that you have named the five most important areas in your life, I am going to ask you to rate how each of these areas are for you at the moment. First I will show you an example of how the rating is done.

First, look at this box (**indicate**). As you can see, there are spaces at the bottom in which I can write the five important areas of my life (**indicate**), and there is a scale along the left hand side (**indicate**). The scale ranges from 'worst possible' on the bottom to 'best possible' on the top, and passes through levels such as **very bad – bad – neither good nor bad – good** and **very good** between the two extremes.

The first important area of my life is **name of cue** and if this is going very well at the moment, I can show this by drawing a bar like this (**draw bar 80mm high**). I am using the scale (**indicate**) to decide how high my bar should be. The nearer I draw the bar to the bottom line, the poorer my rating of that area of my life and the nearer I draw it to the top line, the better my rating of that area of my life. A mark in the middle range would indicate that I am rating life as neither good nor bad, but somewhere in between.

Second cue: if name of cue (**write in the second place**) is going as well as is possible, I would rate it by drawing a bar like this (draw bar 100mm high)

Third cue: if name of cue (**write in the third place**) is going very badly, I would rate it like this (**draw bar 15 mm high**)

Fourth cue: if name of cue (**write in the fourth place**) is just all right, or fifty / fifty, I would rate it like this (draw bar 50mm high)

Fifth cue: name of cue (**write in fifth place**) – (**draw random rating**)

This provides a picture of life as I might think of it at the moment.

Now I want you to rate the five most important areas of your life, as you see presented here (**indicate**). Firstly draw a bar which represents how you would rate yourself on each of these areas at the moment. As in the example I've just shown you, the nearer you draw the bar to the bottom, the poorer you are rating that area of your life and the nearer you draw it to the top line, the better your rating of that area of your life.

Probe: Rate how each of these areas are for you at the moment

Appendix VII Baseline interview schedule (incorporating SEIQoL-DW and EQ-5D)

Probe: What parts of your life are most important?
What things are most important?
The most important things in my life are ... ?

Note: Do not give examples

Family, relationships, health, finances, living conditions, work, social life, leisure activities, religion / spiritual life

Explore their own understanding / definition of each cue (compare and contrast if necessary)

Determining levels

Now that you have named the five most important areas in your life, I am going to ask you to rate how each of these areas are for you at the moment. First I will show you an example of how the rating is done.

First, look at this box (**indicate**). As you can see, there are spaces at the bottom in which I can write the five important areas of my life (**indicate**), and there is a scale along the left hand side (**indicate**). The scale ranges from 'worst possible' on the bottom to 'best possible' on the top, and passes through levels such as **very bad – bad – neither good nor bad – good** and **very good** between the two extremes.

The first important area of my life is **name of cue** and if this is going very well at the moment, I can show this by drawing a bar like this (**draw bar 80mm high**). I am using the scale (**indicate**) to decide how high my bar should be. The nearer I draw the bar to the bottom line, the poorer my rating of that area of my life and the nearer I draw it to the top line, the better my rating of that area of my life. A mark in the middle range would indicate that I am rating life as neither good nor bad, but somewhere in between.

Second cue: if name of cue (**write in the second place**) is going as well as is possible, I would rate it by drawing a bar like this (**draw bar 100mm high**)

Third cue: if name of cue (**write in the third place**) is going very badly, I would rate it like this (**draw bar 15 mm high**)

Fourth cue: if name of cue (**write in the fourth place**) is just all right, or fifty / fifty, I would rate it like this (**draw bar 50mm high**)

Fifth cue: name of cue (**write in fifth place**) – (**draw random rating**)

This provides a picture of life as I might think of it at the moment.

Now I want you to rate the five most important areas of your life, as you see presented here (indicate). Firstly draw a bar which represents how you would rate yourself on each of these areas at the moment. As in the example I've just shown you, the nearer you draw the bar to the bottom, the poorer you are rating that area of your life and the nearer you draw it to the top line, the better your rating of that area of your life.

Probe: Rate how each of these areas are for you at the moment

What is affecting your rating of (each of these) areas?

Probe for knee related causes (if apparent)

Weighting

I would like you to show me how important the five areas of life you have nominated are in relation to each other, by using this disk (indicate). People often value some areas in life as more important than others. This disk allows you to show me how important each area in your life is by giving the more important areas a larger area of the disk, and the less important areas a smaller area of the disk. In my life, for example, name of cue is about this important (30%), Name of cue is less important than name of cue, so it has only this much of the pie (20%). Name of cue (3) on the other hand is more important than name of cue (1), so it has this much of the pie (manipulate to show 40% of space). Finally name of cue 4 and name of cue 5 are the least important areas of life for me, and I value them about the same (5% each).

Now thinking about the five areas of life you have mentioned. I would like you to show me how important these areas are in relation to each other by moving disks around until their relative size represents your view of their importance.

Can you tell me why you have ranked them in this way?

Why is one more important than the other?

How easy is it to rank them in this way?

Direct assessment of content change

If I had asked you **one year ago** to describe the five most important areas that affect your overall quality of life would you have come up with the same list:

Would you have left any of the current list out?

Would you have included different areas?

If so, why?

If no differences, would you have ranked them in a different order?

A knee story

Turning specifically to your knee problem, can you tell me how your knee problem started?

Probes / follow-up questions:

- What do you think caused it?
- What is the nature of the problem?
- How does it affect you now – in what ways?

Daily / long-term / work / social / family / personal / coping / other

- Does it affect any of the areas of life that you have already identified as important?
- How do you think it will affect you in the future?

Note: look for references to change, adaptation, comparison (social and temporal)

Do you think that your knee problem and how it has affected you has changed the way you think about yourself?

Summarise issues from this section / thanks interviewee for their insights

Finally can I ask you what is your current or most recent job?

Cue definitions record form

Description of cue	Cue label
1	_____
2	_____
3	_____
4	_____
5	_____

Tick any cues elicited by reading list to person

Cue levels record form

Best possible

Worst possible

The form consists of a large, empty rectangular box. On the right side of the box, there is a vertical scale with 10 tick marks. The scale is oriented vertically, with 'Best possible' at the bottom and 'Worst possible' at the top. The box is intended for recording cue levels.

This questions ask about your health in general. By placing a cross in one box in each group below, please indicate in which statement best describes your own health state today.

Do not cross more than one box in each group.

1. Mobility

I have no problems in walking about

I have some problems in walking about

I am confined to bed

2. Self-Care

I have no problems with self-care

I have some problems washing or dressing myself

I am unable to wash or dress myself

3. Usual activities (e.g. work, study, housework, family or leisure activities)

I have no problems with performing my usual activities

I have some problems with performing my usual activities

I am unable to perform my usual activities

4. Pain/Discomfort

I have no pain or discomfort

I have moderate pain or discomfort

I have extreme pain or discomfort

5. Anxiety/Depression

I am not anxious or depressed

I am moderately anxious or depressed

I am extremely anxious or depressed

Appendix XIII Follow-up interview schedule (incorporating SEIQoL-DW and EQ-5D)

Follow-up interview schedule Date:

Response shift follow-up interview schedule

i Introduction

- Introduce self
- Review background to study
- Check interviewee happy with recording / transcribing
- General: Duration / check happy / questions

Aims

As before two elements to conversation:

- Review areas of your life that you believe are most important to your overall quality of life
- Tell me about your knee problem and how things have been since previous visit

ii SEIQOL

i) Introduction – cue elicitation

For each of us, happiness and satisfaction in life depends on those parts or areas of life which are important to us. When these important areas are present or are going well, we are generally happy but when they are absent or are going badly we feel worried or unhappy. In other words, these important areas of life determine the quality of our lives. What is considered important varies from person to person. That which is most important to you may not be so important to me or to your husband / wife / children / parents / friends and vice versa.

I am interested in knowing what the most important areas of your life are at the moment. Most of us don't usually spend a lot of time thinking about these things. Indeed, we often only notice that certain things are important when something happens to change them. Sometimes it is easier to identify what is important by thinking about the areas of life that would (or do) cause us most concern when they are missing or are going badly.

What are the five most important areas of your life at present – the things which make your life a relatively happy or sad one at the moment ... the things that you feel determine the quality of your life?

Probe: What parts of your life are most important?
What things are most important?
The most important things in my life are ... ?

Appendix XIII Follow-up interview schedule (incorporating SEIQoL-DW and EQ-5D)

Follow-up interview schedule ID no: Date:

Note: Do not give examples

Family, relationships, health, finances, living conditions, work, social life, leisure activities, religion / spiritual life

Explore their own understanding / definition of each cue (compare and contrast if necessary)

Cue definitions record form

Description of cue	Cue label
1	_____
2	_____
3	_____
4	_____
5	_____

Tick any cues elicited by reading list to person

Appendix XIII Follow-up interview schedule (incorporating SEIQoL-DW and EQ-5D)

Follow-up interview schedule ID no: Date:

ii) Determining levels

Now that you have named the five most important areas in your life, I am going to ask you to rate how each of these areas are for you at the moment. First I will show you an example of how the rating is done.

First, look at this box (*indicate*). As you can see, there are spaces at the bottom in which I can write the five important areas of my life (*indicate*), and there is a scale along the left hand side (*indicate*). The scale ranges from 'worst possible' on the bottom to 'best possible' on the top, and passes through levels such as **very bad – bad – neither good nor bad – good** and **very good** between the two extremes.

The first important area of my life is **name of cue** and if this is going very well at the moment, I can show this by drawing a bar like this (*draw bar 80mm high*). I am using the scale (*indicate*) to decide how high my bar should be. The nearer I draw the bar to the bottom line, the poorer my rating of that area of my life and the nearer I draw it to the top line, the better my rating of that area of my life. A mark in the middle range would indicate that I am rating life as neither good nor bad, but somewhere in between.

Second cue: if name of cue (*write in the second place*) is going as well as is possible, I would rate it by drawing a bar like this (*draw bar 100mm high*)

Third cue: if name of cue (*write in the third place*) is going very badly, I would rate it like this (*draw bar 15 mm high*)

Fourth cue: if name of cue (*write in the fourth place*) is just all right, or fifty / fifty, I would rate it like this (*draw bar 50mm high*)

Fifth cue: name of cue (*write in fifth place*) – (*draw random rating*)

This provides a picture of life as I might think of it at the moment.

Now I want you to rate the five most important areas of your life, as you see presented here (*indicate*). Firstly draw a bar which represents how you would rate yourself on each of these areas at the moment. As in the example I've just shown you, the nearer you draw the bar to the bottom, the poorer you are rating that area of your life and the nearer you draw it to the top line, the better your rating of that area of your life.

Probe: Rate how each of these areas are for you at the moment

Appendix XIII Follow-up interview schedule (incorporating SEIQoL-DW and EQ-5D)

Follow-up interview schedule ID no: Date:

What is affecting your rating of (each of these) areas?

Probe for knee related causes (if apparent)

iii) Weighting

I would like you to show me how important the five areas of life you have nominated are in relation to each other, by using this disk (**indicate**). People often value some areas in life as more important than others. This disk allows you to show me how important each area in your life is by giving the more important areas a larger area of the disk, and the less important areas a smaller area of the disk. In my life, for example, **name of cue** is about this important (30%), **Name of cue** is less important than **name of cue**, so it has only this much of the pie (20%). **Name of cue (3)** on the other hand is more important than **name of cue (1)**, so it has this much of the pie (manipulate to show 40% of space). Finally **name of cue 4** and **name of cue 5** are the least important areas of life for me, and I value them about the same (5% each).

Now thinking about the five areas of life you have mentioned. I would like you to show me how important these areas are in relation to each other by moving disks around until their relative size represents your view of their importance.

Can you tell me why you have ranked them in this way?

Why is one more important than the other?

How easy is it to rank them in this way?

Appendix XIII Follow-up interview schedule (incorporating SEIQoL-DW and EQ-5D)

Follow-up interview schedule ID no: Date:

iii Health transition items

Compared with six months ago how much is your knee interfering with your daily activities now (including work, housework, hobbies and sporting activities)?

- | | | | | | | |
|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Much less | Somewhat less | A little less | About the same | A little more | Somewhat more | Much more |

Compared with six months ago how much of the time do you spend worrying about your knee?

- | | | | | | | |
|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Much less | Somewhat less | A little less | About the same | A little more | Somewhat more | Much more |

Appendix XIII Follow-up interview schedule (incorporating SEIQoL-DW and EQ-5D)

Follow-up interview schedule ID no: Date:

Compared with six months ago, how is your knee now?

Please cross one box	A very great deal worse	<input type="checkbox"/>
	A great deal worse	<input type="checkbox"/>
	A good deal worse	<input type="checkbox"/>
	Moderately worse	<input type="checkbox"/>
	Somewhat worse	<input type="checkbox"/>
	A little worse	<input type="checkbox"/>
	Almost the same, hardly any worse at all	<input type="checkbox"/>
	No change	<input type="checkbox"/>
	Almost the same, hardly any better at all	<input type="checkbox"/>
	A little better	<input type="checkbox"/>
	Somewhat better	<input type="checkbox"/>
	Moderately better	<input type="checkbox"/>
	A good deal better	<input type="checkbox"/>
	A great deal better	<input type="checkbox"/>
	A very great deal better	<input type="checkbox"/>

Appendix XIII Follow-up interview schedule (incorporating SEIQoL-DW and EQ-5D)

Follow-up interview schedule ID no: Date:

iv Narrative description

Turning specifically to your knee problem, can you tell me how your knee has been since we last met?

Probes / follow-up questions:

- In what ways does it affect you **now**?

Daily / long-term / work / social / family / personal / coping / other

- Does it affect any of the areas of life that you have already identified as important?
- How do you think you have coped / adapted to your knee problem?
- Do you understand any more about what caused the problem?
- How do you think it will affect you in the future?

Note: look for references to change, adaptation, comparison (social and temporal)

Do you think that your knee problem and how it has affected you has changed the way you think about yourself or the way that you approach your life?

Appendix XIII Follow-up interview schedule (incorporating SEIQoL-DW and EQ-5D)

Follow-up interview schedule ID no: Date:

v Response shift

At baseline interviewee's cues / rankings were:

Rank	Name of cue	Description	Weighting
1			
2			
3			
4			
5			

Either ask 1, 2 or 3 below:

1a) If there are **differences in cue content** (name or description):

When I asked you to describe important areas of your life earlier you mentioned **name of cue** which we didn't discuss in our first meeting.

Why do you think you have mentioned this / these **now**?

1b)

Similarly, in our first meeting you mentioned **name of cue** but not now - why do you think that you haven't mentioned that **this** time?

2) If there are **differences in cue ranking** (name or description):

Either: A (before) highest ranked item moving down.
B Large change in ranking of item (e.g. from 5

to 2)

A Earlier, when I asked you to put the five areas of your life in order of importance, you put **name of cue** at the top. In our first meeting you put **name of cue** at the top. Why do you think you have changed the order you have placed this item?

B Earlier, when I asked you to put the five areas of your life in order of importance, you put name of cue in [] place. In our first meeting you put the same area in [] place. Why do you think you have changed the order of this item?

3) If there are no differences in cue content or ranking

Do you think that you have changed in what you would consider to an important area of life since we first spoke six months ago?

Appendix XIII Follow-up interview schedule (incorporating SEIQoL-DW and EQ-5D)

Follow-up interview schedule ID no: Date:

LOT-R

Please be as honest and accurate as you can throughout. Try not to let your response to one statement influence your responses to other statements. There are no "correct" or "incorrect" answers. Answer according to your own feelings, rather than how you think "most people" would answer.

- A = I agree a lot
- B = I agree a little
- C = I neither agree nor disagree
- D = I disagree a little
- E = I disagree a lot

- 1 In uncertain times, I usually expect the best.
- 2 It's easy for me to relax.
- 3 If something can go wrong for me, it will.
- 4 I'm always optimistic about my future
- 5 I enjoy my friends a lot
- 6 It's important for me to keep busy
- 7 I hardly ever expect things to go my way
- 8 I do't get upset too easily
- 9 I rarely count on good things happening to me
- 10 Overall, I expect more good things to happen to me than bad

Appendix XIII Follow-up interview schedule (incorporating SEIQoL-DW and EQ-5D)

Follow-up interview schedule ID no: Date:

These questions ask about your health in general. By placing a cross in one box in each group below, please indicate in which statement best describes your own health state today.

Do not cross more than one box in each group.

1. Mobility

I have no problems in walking about

I have some problems in walking about

I am confined to bed

2. Self-Care

I have no problems with self-care

I have some problems washing or dressing myself

I am unable to wash or dress myself

3. Usual activities (e.g. work, study, housework, family or leisure activities)

I have no problems with performing my usual activities

I have some problems with performing my usual activities

I am unable to perform my usual activities

4. Pain/Discomfort

I have no pain or discomfort

I have moderate pain or discomfort

I have extreme pain or discomfort

5. Anxiety/Depression

I am not anxious or depressed

I am moderately anxious or depressed

I am extremely anxious or depressed

Appendix XIII Follow-up interview schedule (incorporating SEIQoL-DW and EQ-5D)

Follow-up interview schedule ID no: Date:

Finally, I would like you to think back to our first interview. **Could you please answer the same questions about your health in general.** By placing a cross in one box in each group below, please indicate in which statement best describes your own health state **as it was six months ago.**

Do not cross more than one box in each group.

1. Mobility

I have no problems in walking about

I have some problems in walking about

I am confined to bed

2. Self-Care

I have no problems with self-care

I have some problems washing or dressing myself

I am unable to wash or dress myself

3. Usual activities (e.g. work, study, housework, family or leisure activities)

I have no problems with performing my usual activities

I have some problems with performing my usual activities

I am unable to perform my usual activities

4. Pain/Discomfort

I have no pain or discomfort

I have moderate pain or discomfort

I have extreme pain or discomfort

5. Anxiety/Depression

I am not anxious or depressed

I am moderately anxious or depressed

I am extremely anxious or depressed

Appendix XIII Follow-up interview schedule (incorporating SEIQoL-DW and EQ-5D)

Follow-up interview schedule ID no: Date:

Cue levels record form

Best possible



Worst possible

Appendix IX Individual case summaries

Brief case summaries are provided for all study patients and each begins with details of **study and demographic variables**, including the patient's study identifier, age, gender, stratification status (arthroscopy or recently referred) and their occupation (if available / relevant). A brief **clinical summary** is provided. This includes firstly their state at enrolment according to either their referral letter or arthroscopy diary notes. The former varied in level of detail although most contained details of presenting symptoms or signs. The latter was usually very brief and may simply have indicated (for example) that a ligament or meniscus was involved.

The summary also includes the patient's narrative of their knee-related history. This is their view of the knee problem up to the point of the baseline interview. Given that patients sometimes were unaware of their diagnosis and even the purpose of either received, or planned treatment, these should not be taken as necessarily medically valid descriptions. However, it does represent the patient's perspective upon their problem and received care. Finally, the patients' account of their knee problem from the follow-up interview is included. For arthroscopy patients this may include details of the operation, and for all patients may include description of the progress of the knee complaint. Relevant or significant co-morbidity is also described as appropriate. As many of the issues above are described in other areas of the thesis, the clinical summaries are have been kept as brief as possible.

The five **SEIQoL cues** nominated by the patient at baseline and follow-up interviews are visually displayed showing their reported level and weight (using histograms and pie charts respectively). The displayed cue labels are generally those used by the patients.

Summary quantitative data from the baseline and follow-up interviews is provided for SEIQoL-DW and for EQ-5D. The latter incorporates the conventional pretest and posttest scores and the retrospective thentest scores. Finally, the summary Life Orientation Test (Revised) score of dispositional optimism is provided (scores may range from 0 to 24).

Patient #1

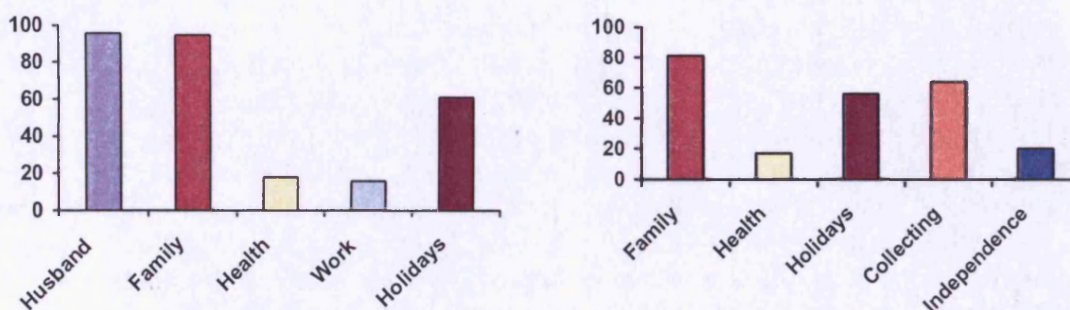
Age: 41 **Gender:** Female **Status:** Arthroscopy **Job:** Retail manager

The patient described a long history of knee problems originating from a childhood injury and subsequent intermittent intervention including removal of the kneecap. She has attended a pain clinic due to her knee. Her expectation and desire is for a partial knee-replacement (but has been advised that she is too young). Following her arthroscopy, the patient has been left more debilitated. She has been advised that the next step would be to remove some cartilage material which would be grown 'in vitro' and re-implanted. However, this is finance dependent.

SEIQoL cues, levels and weights

Baseline (2/12/02) **Follow-up (8/7/03)**

Levels



Weights



Summary quantitative data

Assessment	Baseline	Follow-up
SEIQoL (total)	65.35	50.34
EQ-5D	Pretest: 0.19	Con. Posttest: 0.19 Thentest: 0.52
Life Orientation Test	9.0	

Appendix IX Individual case summaries

Patient #2

Age: 40 **Gender:** Male **Status:** Arthroscopy **Job:** Civil engineer

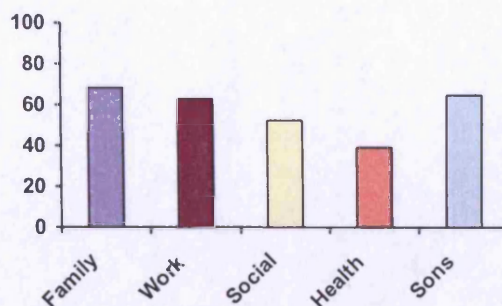
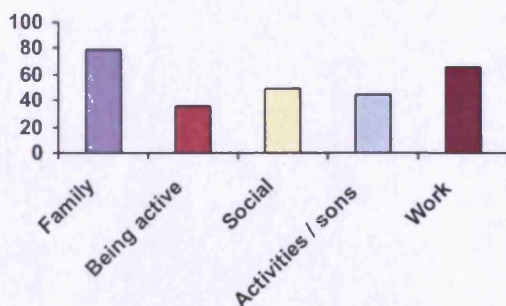
Summary: The patient described a long history of knee problems following an injury 5-6 years earlier and previous surgery. At baseline he was listed for revision to the anterior cruciate ligament. At follow-up he reported deterioration in knee function – although stronger, he described reduced range of movement in the joint. This has (for example) stopped him being able to run, which he was previously able to do.

SEIQoL cues, levels and weights

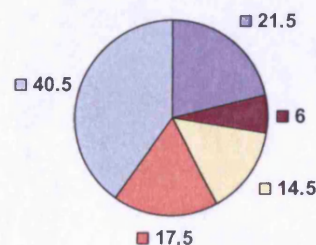
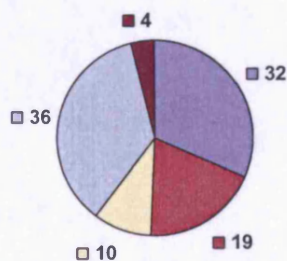
Baseline (6/12/02)

Follow-up (20/6/03)

Levels



Weights



Summary quantitative data

Assessment	Baseline	Follow-up
SEIQoL (total)	55.37	57.86
EQ-5D	Pretest: 0.73	Con. Posttest: 0.62 Thentest: 1.00
Life Orientation Test	12	

Patient #3

Age: 36 **Gender:** Female **Status:** Arthroscopy **Job:** Medical doctor

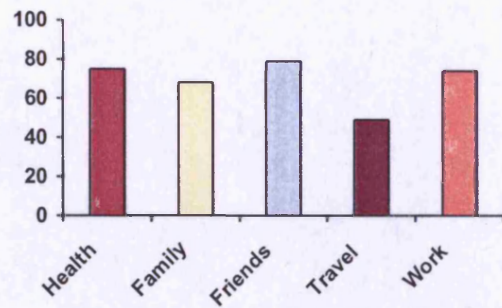
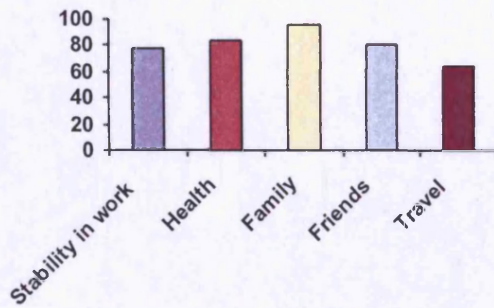
Summary: At baseline, the patient was listed for arthroscopy. She described having an injury whilst exercising 4-5 years previously and subsequent instability. She had already undergone previous surgery on her knee and considered that the forthcoming surgery would be aimed at identifying what continues to be wrong. However, at follow-up she reported that she actually received a new ligament.

SEIQoL cues, levels and weights

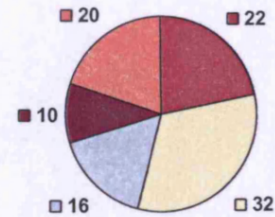
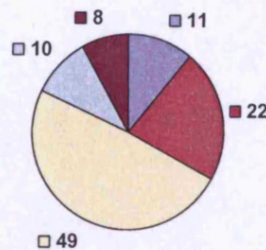
Baseline (6/12/02)

Follow-up (15/7/03)

Levels



Weights



Summary quantitative data

Assessment	Baseline	Follow-up
SEIQoL (total)	87.32	70.6
EQ-5D	Pretest: 0.76	Con. Posttest: 1.00 Thentest: 0.88
Life Orientation Test	17	

Appendix IX Individual case summaries

Patient #4

Age: 24 **Gender:** Female **Status:** Arthroscopy **Job:** Student

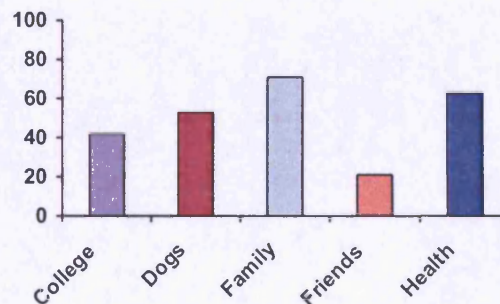
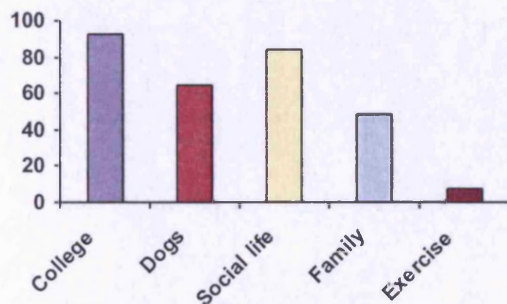
Summary: At baseline the patient was listed for an operation due to an absent anterior cruciate ligament. She described a long standing problem with her knee emanating from an injury she suffered at school, about eight years previously. She had originally been told that there had been a dislocation and her leg had been placed in plaster. Subsequently, she had her leg plastered another 10 or more times over several years (each time for about six weeks). She has seen several secondary care doctors and most recently had an initial operation to tighten her knee ligaments. Her arthroscopy was intended to be the first of two to continue this process. At follow-up, the patient reported that the operation had been to re-construct the cruciate ligament. She has had subsequent physiotherapy, but at the time of follow-up interview, reported that she could not straighten her leg.

SEIQoL cues, levels and weights

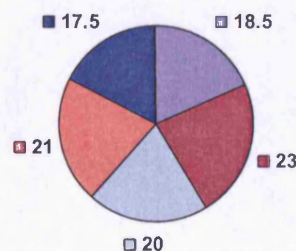
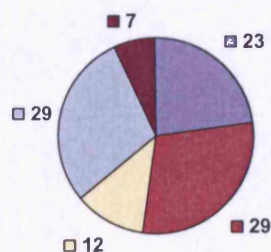
Baseline (9/12/02)

Follow-up (26/6/03)

Levels



Weights



Summary quantitative data

Assessment	Baseline	Follow-up
SEIQoL (total)	64.57	63.67
EQ-5D	Pretest: 0.69	Con. Posttest: 0.69 Thentest: 0.69
Life Orientation Test	14	

Appendix IX Individual case summaries

Patient #5

Age: 52 **Gender:** Male **Status:** Arthroscopy **Job:** Electrician

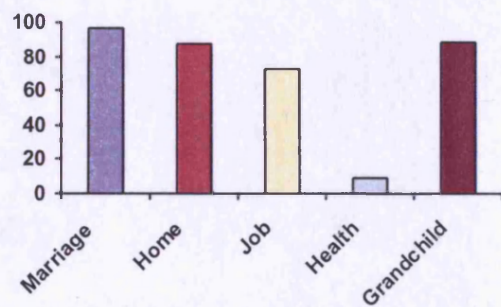
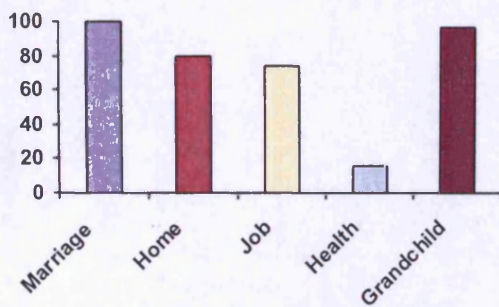
Summary: The patient was listed at baseline for arthroscopy. Although he described a number of other medical complaints (and a history of operations, for example, on his back), he is most debilitated by his knees. Concern about his deteriorating physical health has also led him to be depressed and at baseline he was on anti-depressant medication. Previously very physically active, he has gradually noticed problems with both knees and has already had an operation on his right knee. Now his left knee has been causing similar problems over the last five years. At follow-up interview he reported that he actually had arthroscopies on each knees and has been told he has arthritis in both. He reported much less pain and greater mobility as a consequence of the operations.

SEIQoL cues, levels and weights

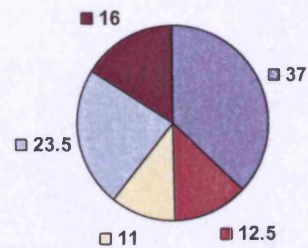
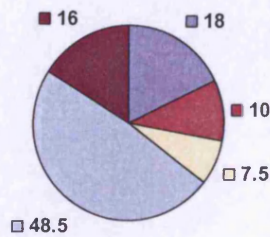
Baseline (12/12/02)

Follow-up (30/6/03)

Levels



Weights



Summary quantitative data

Assessment	Baseline	Follow-up
SEIQoL (total)	54.73	71.28
EQ-5D	<i>Pretest: 0.02</i>	<i>Con. Posttest:0.52 Thentest:0.02</i>
Life Orientation Test	13	

Appendix IX Individual case summaries

Patient #6

Age: 19 **Gender:** Female **Status:** Referred **Job:** Call centre operator

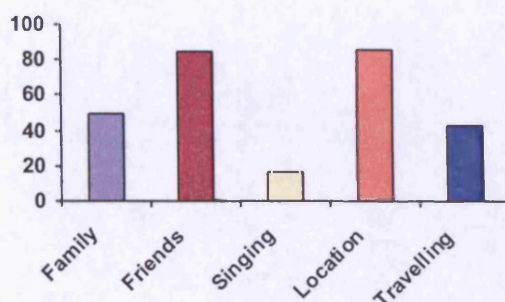
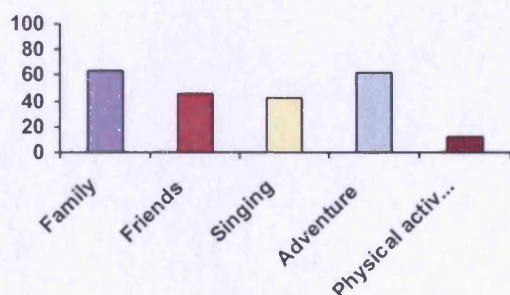
Summary: At baseline the patient had been referred to orthopaedics with pain, instability, swelling, hotness of the knee and clicking. She described a history of problems going back several years, but with a recent deterioration within the last few months. She reported intermittent periods of pain and swelling which had become more frequent. At baseline interview she had already been seen by the surgeon (despite only a recent referral) and had been sent for MRI. At follow-up she reported a continuation of her problems which, despite some improvement are causing her difficulties in work and other aspects of her life. She was still waiting for the result of a scan two months previously, about which she is getting frustrated.

SEIQoL cues, levels and weights

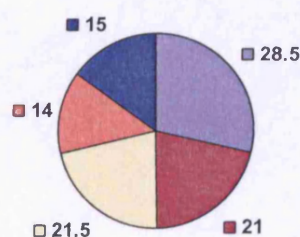
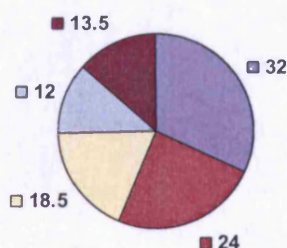
Baseline (16/12/02)

Follow-up (1/8/03)

Levels



Weights



Summary quantitative data

Assessment	Baseline	Follow-up
SEIQoL (total)	48.35	54.25
EQ-5D	Pretest: 0.69	Con. Posttest: 0.69 Thentest: 0.62
Life Orientation Test	21	

Appendix IX Individual case summaries

Patient #7

Age: 53 **Gender:** Male **Status:** Referred **Job:** Local government administrator

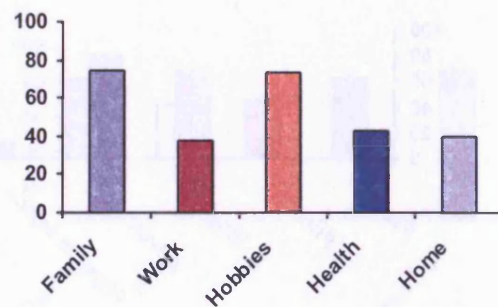
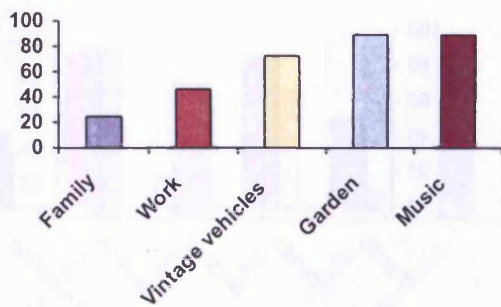
Summary: At baseline, orthopaedic referral information included references to meniscus and knee pain. The patient reported a long history (about nine years) of problems including locking of the knee. He has previously had an MRI and a torn cartilage diagnosed. However, an operation was deferred because of his age and he described how his case really didn't get pursued – until recently when it was causing him more problems. At follow-up he is on an orthopaedic waiting list and his knee may have deteriorated but he appeared less concerned about it. However, he has concerns about other health problems.

SEIQoL cues, levels and weights

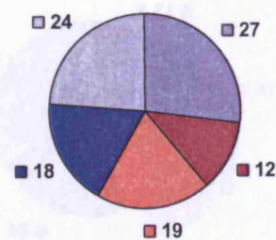
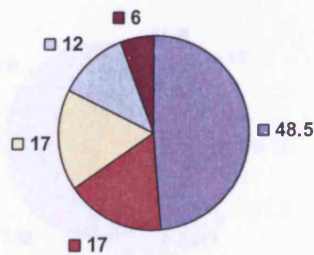
Baseline (17/12/02)

Follow-up (4/8/03)

Levels



Weights



Summary quantitative data

Assessment	Baseline	Follow-up
SEIQoL (total)	47.6	55.75
EQ-5D	Pretest: 0.69	Con. Posttest: 0.88 Thentest: 0.69
Life Orientation Test	13	

Appendix IX Individual case summaries

Patient #8

Age: 39 **Gender:** Male **Status:** Referred **Job:** Chef

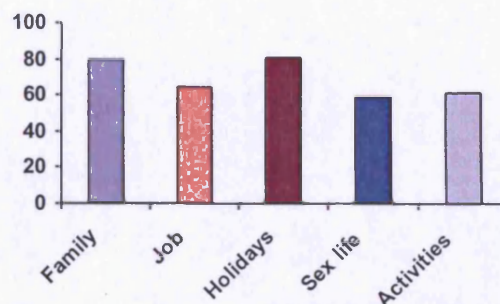
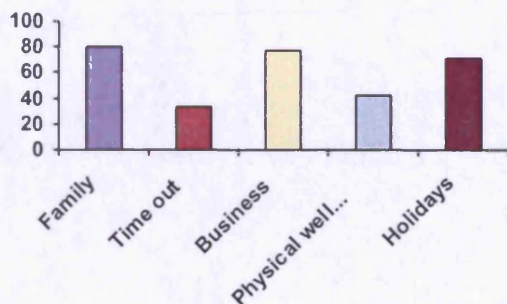
Summary: At baseline his referral to orthopaedics indicated a meniscal problem with pain and swelling. The patient described injuring his knee many years previously whilst playing sport but not suffering significant problems in the interim. He was concerned about his ability to conduct his work but has already had physiotherapy and reflexology. He reported that most of his difficulties appear to have subsided (although he still described the cruciate ligament as 'loose'). He was reluctant to have surgery and also was unhappy to take pain-killers. At follow-up his knee remained improved with only some aching. He suggested that he had modified his activities to accommodate his knee problem.

SEIQoL cues, levels and weights

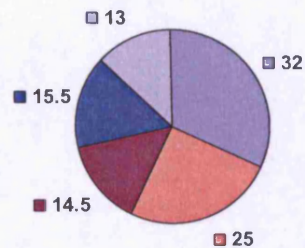
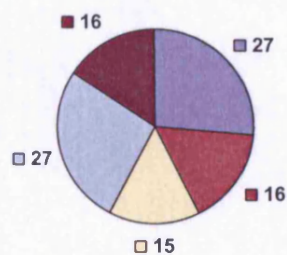
Baseline (20/12/02)

Follow-up (9/7/03)

Levels



Weights



Summary quantitative data

Assessment	Baseline	Follow-up
SEIQoL (total)	60.67	70.42
EQ-5D	Pretest: -	Con. Posttest: 1.0 Thentest: 0.29
Life Orientation Test	13	

Appendix IX Individual case summaries

Patient #9

Age: 51 **Gender:** Male **Status:** Arthroscopy **Job:** Self-employed businessman

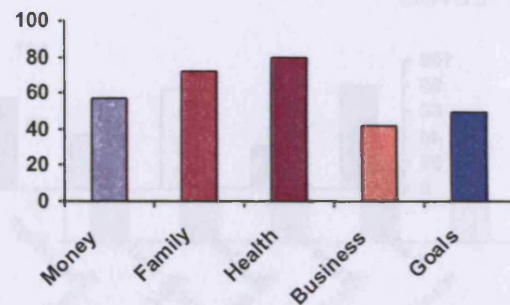
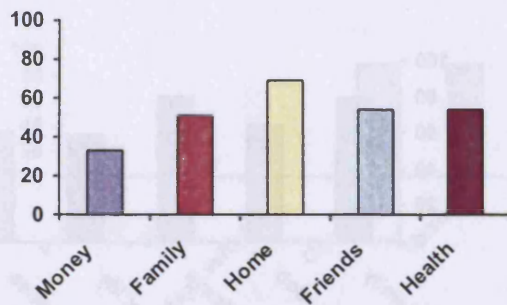
Summary: At baseline the patient had been listed for partial lateral meniscectomy. He had accessed the orthopaedic service via a special knee injury clinic (self-referral). He described an active sporting past and a gradual deterioration in knee function, although he continued to play sport (e.g. squash). He reported that he has had previous surgery on his knee. His main response seems to be irritation at clicking in the joint. Following surgery, he reports that his knee has recovered well and he has even returned to skiing.

SEIQoL cues, levels and weights

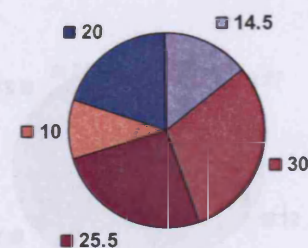
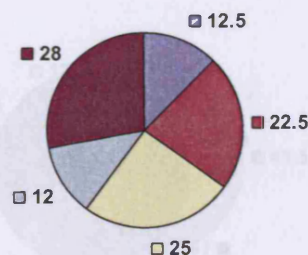
Baseline (20/12/02)

Follow-up (29/7/03)

Levels



Weights



Summary quantitative data

Assessment	Baseline	Follow-up
SEIQoL (total)	54.45	64.27
EQ-5D	Pretest: 0.69	Con. Posttest: 1.0 Thentest: 0.80
Life Orientation Test	16	

Appendix IX Individual case summaries

Patient #10

Age: 51 **Gender:** Male **Status:** Referred **Job:** Retail worker

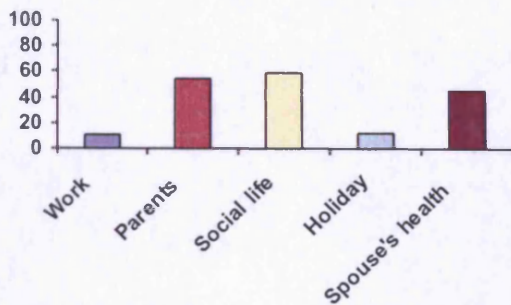
Summary: The patient's referral letter described the meniscus, pain and swelling. At baseline, he described a recent history of clicking and locking especially at work where he is a manual worker in a furniture retail store. He reported having had an x-ray and it revealing cartilage damage and bone degeneration. At follow-up he is still waiting for an orthopaedic appointment and his physical state had not improved. Furthermore, his social situation had worsened (he was off work, had moved in with his elderly parents to care for them and his wife had been in hospital for the previous four weeks with alcohol-related liver problems).

SEIQoL cues, levels and weights

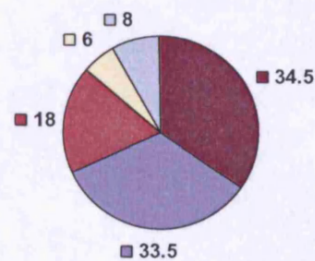
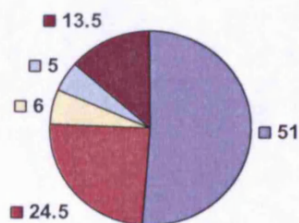
Baseline (7/1/03)

Follow-up (11/7/03)

Levels



Weights



Summary quantitative data

Assessment	Baseline	Follow-up
SEIQoL (total)	28.93	26.31
EQ-5D	<i>Pretest: 0.10</i>	<i>Con. Posttest: 0.03 Thentest: 0.69</i>
Life Orientation Test	18	

Patient #11

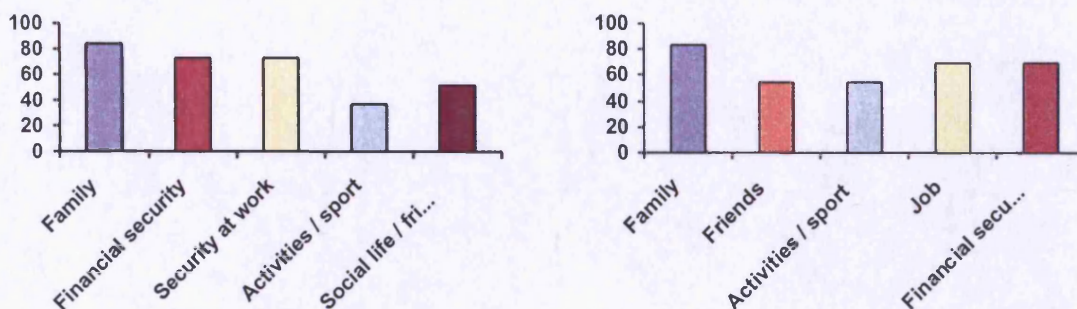
Age: 44 **Gender:** Male **Status:** Referred **Job:** Teacher

Summary: The original referral letter described a meniscus problem following an injury, with pain and swelling. The patient injured his knee whilst playing football and initially had ligament damage diagnosed. However, he continued various physical activities but gradually developed stiffness and swelling which prompted him to consult his general practitioner. At follow-up he was still waiting for an orthopaedic appointment, although he was not substantially troubled by his knee. He had made some adaptations to account for his physical limitations.

SEIQoL cues, levels and weights

Baseline (8/1/03) **Follow-up (5/8/03)**

Levels



Weights



Summary quantitative data

Assessment	Baseline	Follow-up
SEIQoL (total)	71.14	72.08
EQ-5D	Pretest: 0.69	Con. Posttest: 1.0 Thentest: 0.76
Life Orientation Test	8	

Patient #12

Age: 29 **Gender:** Female **Status:** Arthroscopy **Job:** Customer relations officer

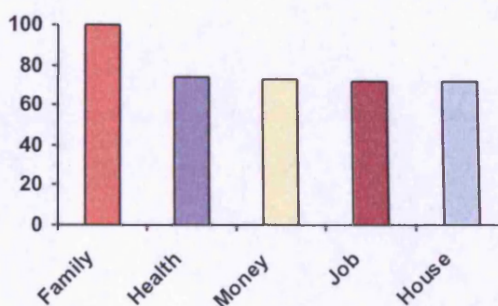
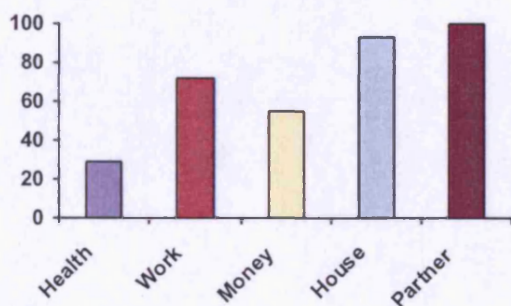
Summary: The patient was listed at baseline for anterior ligament reconstruction. She described at baseline a long history of problems with the knee (approximately 14 years) with an uncertain origin – although she speculated about sport. She reported problems with pain, the knee giving way and clicking. She also reported developing problems in her other knee. She was unhappy with the wait for investigation and treatment. She was originally seen by her own general practitioner five years ago, she waited two and a half years before seeing a specialist and then having an MRI. Her MRI revealed an absent ligament, hence the planned reconstructive surgery. Subsequent to her operation she is very happy with her recovery and restoration of function.

SEIQoL cues, levels and weights

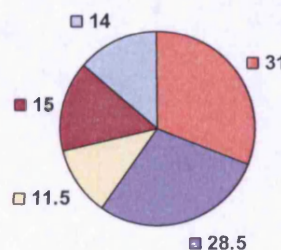
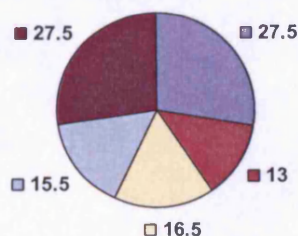
Baseline (8/1/03)

Follow-up (12/8/03)

Levels



Weights



Summary quantitative data

Assessment	Baseline	Follow-up
SEIQoL (total)	60.08	81.36
EQ-5D	Pretest: 0.62	Con. Posttest: 0.88 Thentest: 0.36
Life Orientation Test	17	

Appendix IX Individual case summaries

Patient #13

Age: 32 **Gender:** Female **Status:** Referred **Job:** Retail customer care

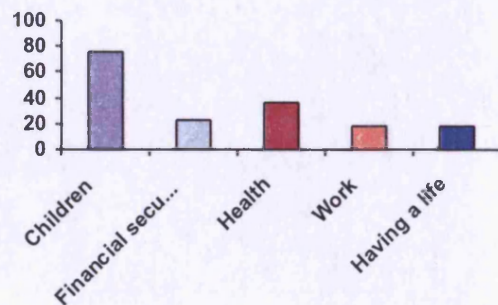
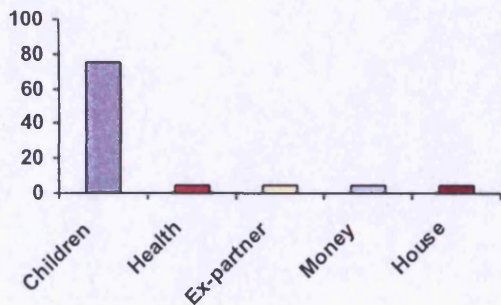
Summary: The patient's baseline referral letter made reference to her meniscus and ligament as well as symptoms of pain, instability and swelling. She described the onset of her knee problem due to a traffic accident whilst a passenger on a bus four years previously. Diagnosis at the time (following a private consultation) included a chipped bone and damaged tendons. Her problem had not resolved and one consequence for her has been substantial weight gain for which she was also seeking treatment (for example, she mentioned gastric bypass). At follow-up she was still waiting for an orthopaedic appointment. She reported worse pain (but better control with analgesia) and stability in other symptoms.

SEIQoL cues, levels and weights

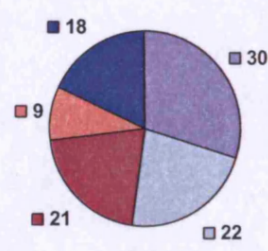
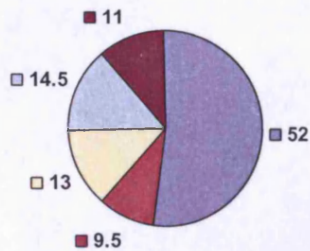
Baseline (9/1/03)

Follow-up (7/8/03)

Levels



Weights



Summary quantitative data

Assessment	Baseline	Follow-up
SEIQoL (total)	41.02	40.28
EQ-5D	Pretest: 0.29	Con. Posttest: 0.08 Thentest: 0.02
Life Orientation Test	2	

Patient #14

Age: 37 **Gender:** Male **Status:** Arthroscopy **Job:** Civil engineer

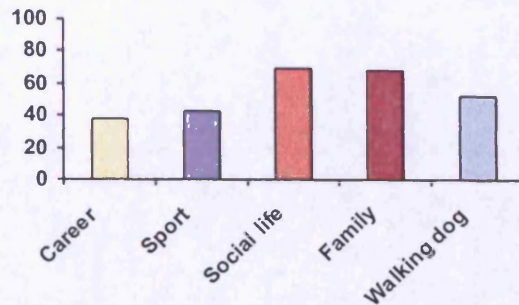
Summary: The patient was listed at baseline for an arthroscopy on his anterior cruciate ligament. At baseline he reported a four year history of problems subsequent to a football injury. After being assessed initially at casualty a few days later, he was referred to orthopaedics where ligament damage was diagnosed. He reported experiencing pain and occasional instability. He had already had an MRI and arthroscopy, but the latter had not repaired the ligament. Instead of going ahead with the planned operation the consultant subsequently decided to monitor and review his case. At follow-up, the patient reports no significant change in his condition and is mainly concerned about the longer-term risk of arthritis.

SEIQoL cues, levels and weights

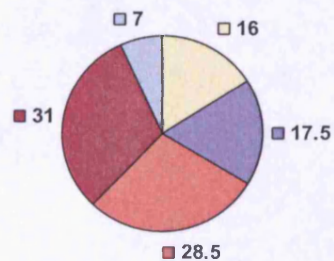
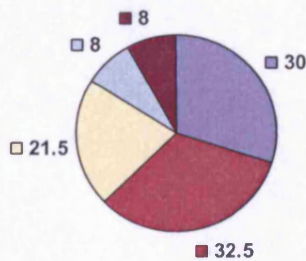
Baseline (14/1/03)

Follow-up (19/8/03)

Levels



Weights



Summary quantitative data

Assessment	Baseline	Follow-up
SEIQoL (total)	63.53	58.06
EQ-5D	<i>Pretest: 0.76</i>	<i>Con. Posttest: 1.0 Thentest: 0.88</i>
Life Orientation Test	14	

Patient #15

Age: 19 **Gender:** Female **Status:** Referred **Job:** Student

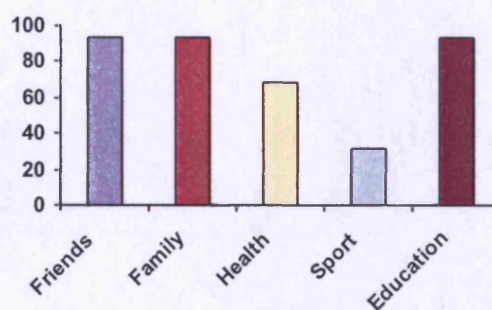
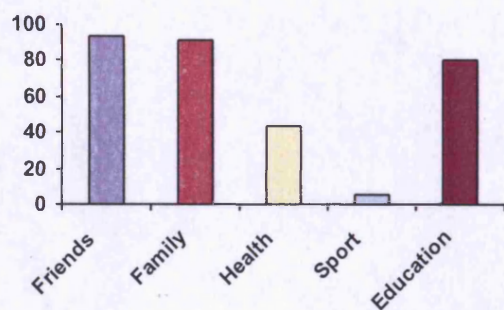
Summary: Although the patient was referred to orthopaedics, clear clinical details were not apparent. She reported at baseline that she had injured her knee four years previously playing netball. She was referred to orthopaedics following a private physiotherapy consultation with the suggestion that she had a torn cartilage. She subsequently had an MRI, arthrogram and finally an arthroscopy to clean out the joint. After initial improvement, it deteriorated again and it has now curtailed her sporting activities. At follow-up she considered her knee to have worsened and she felt that, for example, it was affecting her studies. She was still waiting for her out-patient appointment.

SEIQoL cues, levels and weights

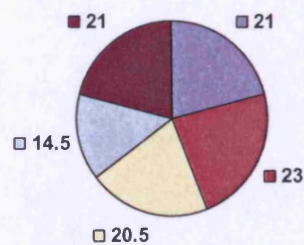
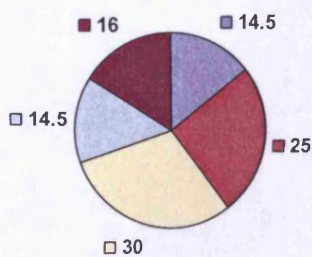
Baseline (15/1/03)

Follow-up (10/9/03)

Levels



Weights



Summary quantitative data

Assessment	Baseline	Follow-up
SEIQoL (total)	62.66	79.10
EQ-5D	<i>Pretest: 0.76</i>	<i>Con. Posttest: 0.36 Thentest: 0.76</i>
Life Orientation Test	21	

Appendix IX Individual case summaries

Patient #16

Age: 28 **Gender:** Male **Status:** Referral **Job:** Call-centre operative

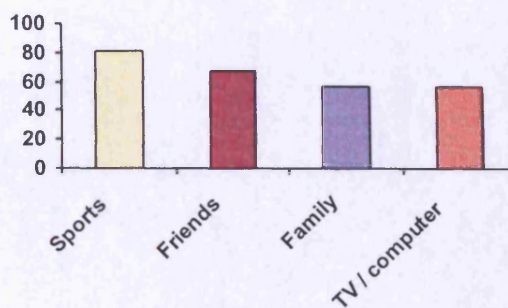
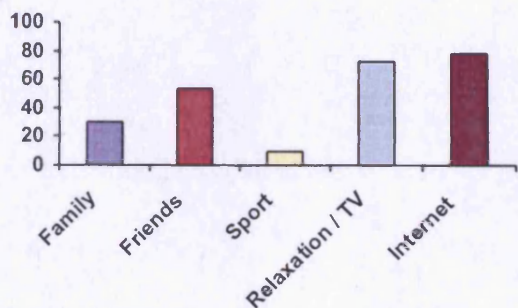
Summary: The patient's referral made reference to the meniscus and symptoms of pain, locking and swelling. At baseline he described a long history of problems with his knee (nearly 20 years) starting in junior school. Following surgery in his early teenage years, he was told that his cartilage was misshapen and it was subsequently removed. It became problematic again in his late teens. He now had a mobility problem; an inability to straighten his knee properly; pain and clicking. At baseline he expressed pessimism about the long-term outlook for his knee. At follow-up he was still awaiting an appointment and reported that he had resumed playing sport. He reported increased worry about his knee at follow-up, and also mentioned concerns about how he might be affected in older age. His return to sport was prompted by his intention to not let his knee problem interfere with his lifestyle.

SEIQoL cues, levels and weights

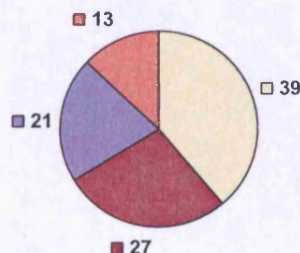
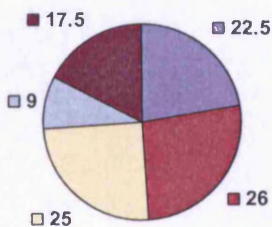
Baseline (22/1/03)

Follow-up (8/9/03)

Levels



Weights



Summary quantitative data

Assessment	Baseline	Follow-up
SEIQoL (total)	43.25	69.06
EQ-5D	Pretest: 0.62	Con. Posttest: 0.81 Thentest: 0.81
Life Orientation Test	9	

Appendix IX Individual case summaries

Patient #17

Age: 51 **Gender:** Female **Status:** Arthroscopy **Job:** School support worker

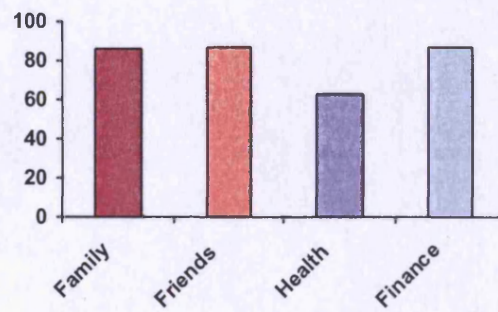
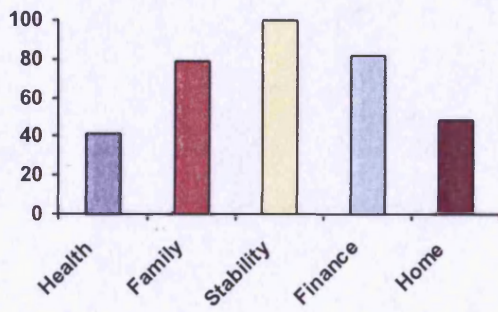
Summary: The patient was listed at baseline for a 'therapeutic endoscopy'. At baseline she reported a gradual onset about three years previously which had led to mobility difficulties around the house and problems, for example, when playing with her young daughter. An operation was initially indicated but she decided not to proceed as her symptoms had improved (she had been taking cod liver oil and undergoing physiotherapy). However, her symptoms returned and she has now been waiting 18 months for her operation. At follow-up she reported an initially slow recuperation from the operation, but ultimately a successful restoration of function. She reported satisfaction that she had it done.

SEIQoL cues, levels and weights

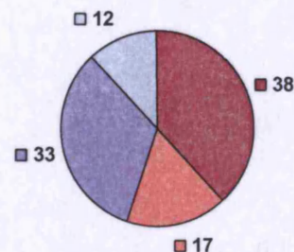
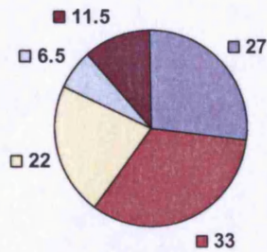
Baseline (22/1/03)

Follow-up (15/9/03)

Levels



Weights



Summary quantitative data

Assessment	Baseline	Follow-up
SEIQoL (total)	69.99	78.70
EQ-5D	Pretest: 0.62	Con. Posttest: 0.81 Thentest: 0.62
Life Orientation Test	16	

Appendix IX Individual case summaries

Patient #18

Age: 64 Gender: Female Status: Referral Job: Retired

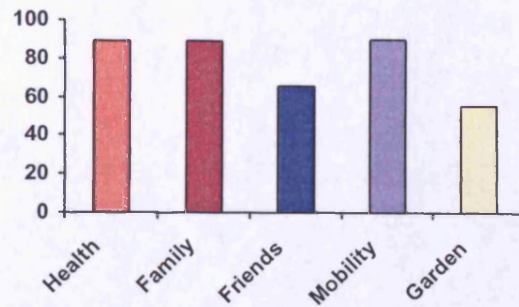
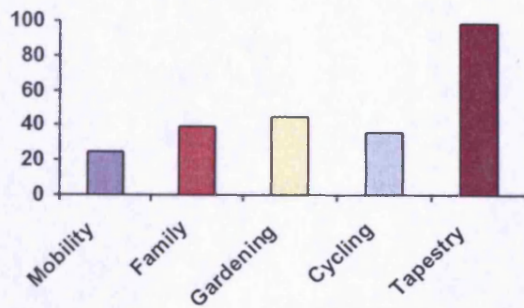
Summary: The patient's referral letter mentioned her meniscus and knee pain. At baseline she reported a gradual onset of symptoms which she initially regarded as arthritis. After consulting her general practitioner, she had an x-ray and she reported having cartilage damage. She started taking anti-inflammatory tablets and paracetamol which significantly reduced the pain, although she still reported some problems. Overall, she was concerned about the potential impact upon her mobility and independence. At follow-up she considered her knee problem to have resolved itself without further intervention and she has resumed her previous activities.

SEIQoL cues, levels and weights

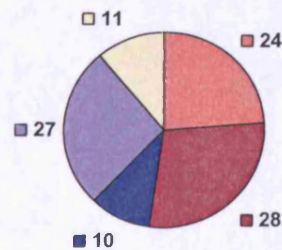
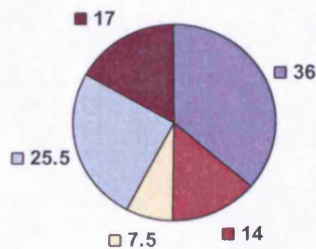
Baseline (22/1/03)

Follow-up (21/8/03)

Levels



Weights



Summary quantitative data

Assessment	Baseline	Follow-up
SEIQoL (total)	43.85	83.34
EQ-5D	Pretest: 0.62	Con. Posttest: 0.85 Thentest: 0.13
Life Orientation Test	14	

Patient #19

Age: 24 **Gender:** Male **Status:** Arthroscopy **Job:** Student

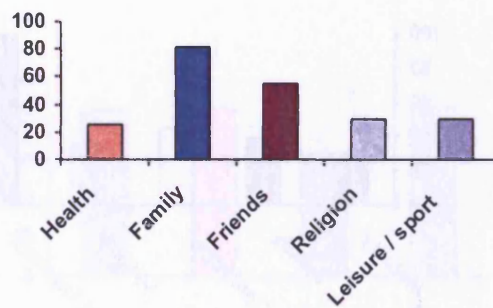
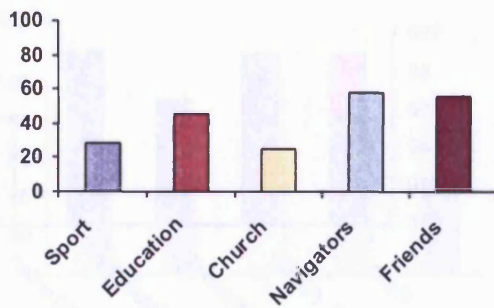
Summary: The patient had been listed for arthroscopy and at the baseline interview had already just had surgery. At baseline he described injuring his knee whilst playing football two years previously and subsequently experiencing pain; swelling; stiffness and locking. He had particular problems when attempting sideways movement – and noted an impact upon his sport and studies. His recollection from the time of the arthroscopy was that there was no major problem with the knee and the joint was cleaned out. At follow-up his knee was better – but had yet to test it substantially by playing sport.

SEIQoL cues, levels and weights

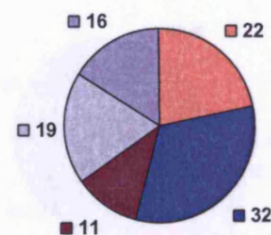
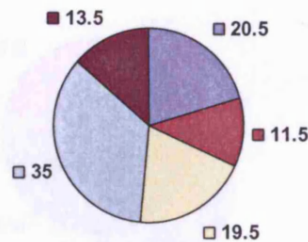
Baseline (24/1/03)

Follow-up (10/9/03)

Levels



Weights



Summary quantitative data

Assessment	Baseline	Follow-up
SEIQoL (total)	43.65	47.94
EQ-5D	Pretest: 0.02	Con. Posttest: 1.0 Thentest: 0.69
Life Orientation Test	16	

Appendix IX Individual case summaries

Patient #20

Age: 29 **Gender:** Male **Status:** Arthroscopy **Job:** Security guard

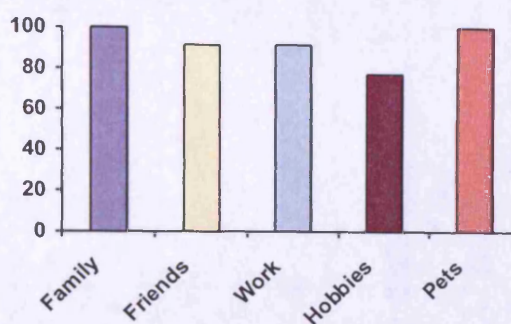
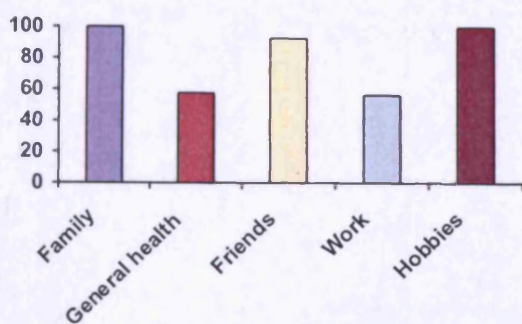
Summary: The patient was listed at baseline for arthroscopy for the removal of a loose body. At baseline he reported being injured in a car crash a number of years previously. However, it was not clear that his knee problem was related to this and he thought the onset of his symptoms was about three years previously. He was a bit annoyed by the time taken to diagnose his complaint (he has had x-rays, an arthrogram and MRI). He reported locking of the knee and pain. Following arthroscopy he felt that his knee problem has substantially resolved but still retained mild concerns about the future possibility of a recurrence.

SEIQoL cues, levels and weights

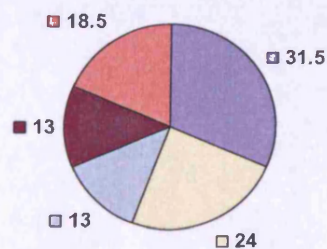
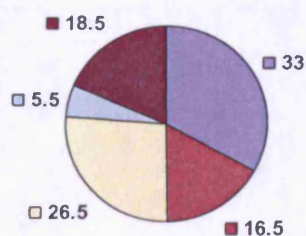
Baseline (24/1/03)

Follow-up (22/8/03)

Levels



Weights



Summary quantitative data

Quantitative Assessment	Baseline	Follow-up
SEIQoL (total)	88.63	93.68
EQ-5D	<i>Pretest: 0.62</i>	<i>Con. Posttest: 0.82 Thentest: 0.52</i>
Life Orientation Test	17	

Patient #21

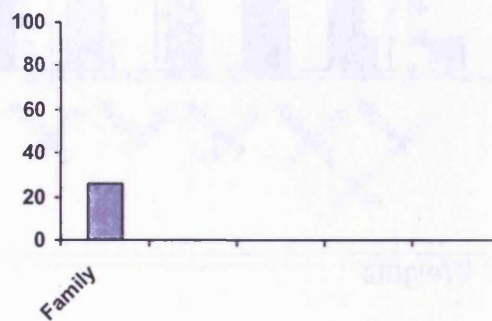
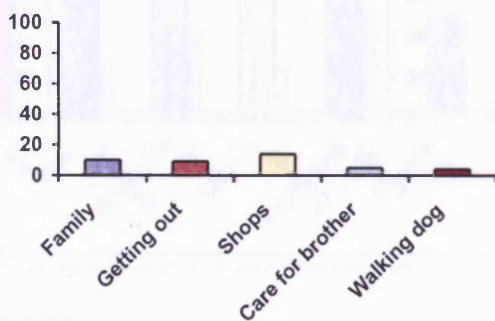
Age: 49 **Gender:** Female **Status:** Arthroscopy **Job:** Housewife / carer

Summary: At baseline the patient was listed for arthroscopy. The patient was interviewed with her son occasionally present during the baseline interview, and husband continually present during the follow-up interview. Her knee was causing her pain and had impacted upon many areas of her life including general mobility, personal care, family activities, weight gain and mental health. Following her operation she felt that there had not been an improvement although she put this down to problems with her hip (she had been told that she was too young to have a hip replacement). She was critical of the care that she received whilst in hospital. She reported difficulties recovering from the operation and appeared somewhat resentful that her progress did not appear to match that of other patients. She was taking medication for depression and redutil to enable weight reduction. She also had asthma but smoked throughout the follow-up interview.

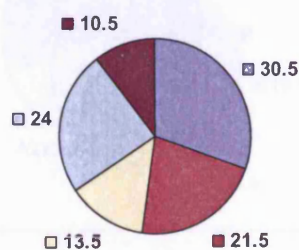
SEIQoL cues, levels and weights

Baseline (28/1/03) **Follow-up (4/9/03)**

Levels



Weights



Summary quantitative data

Assessment	Baseline	Follow-up
SEIQoL (total)	8.49	-
EQ-5D	Pretest: 0.19	Con. Posttest: -0.07 Thentest: -0.07
Life Orientation Test	14	

Appendix IX Individual case summaries

Patient #22

Age: 34 **Gender:** Male **Status:** Referred **Job:** Self-employed carpenter

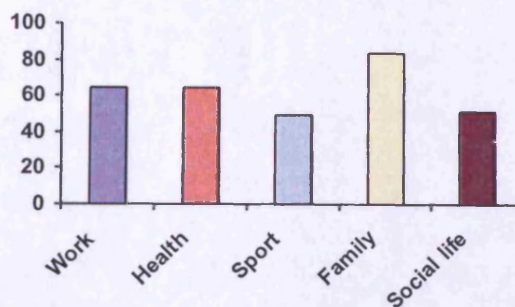
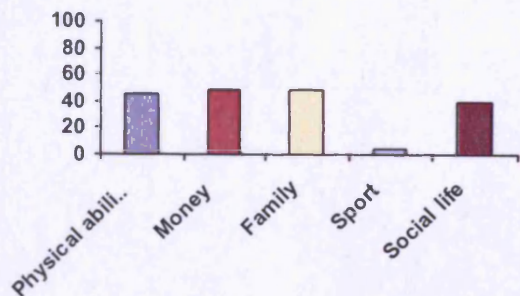
Summary: The patient had been referred at baseline, with references in the referral to meniscus and pain. At baseline the patient described a long-standing problem with his knee (including giving way and clicking) starting not long after he left the army about seven years previously. He remained however, active and subsequently experienced it giving way and being painful and swollen when out running. He thought that his general practitioner had mentioned a chipped piece of bone and cartilage but he was unsure. At follow-up he reported that the knee was no better and, in the absence of further intervention he had made several changes to his work and broader life to accommodate his knee problem.

SEIQoL cues, levels and weights

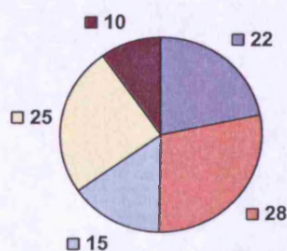
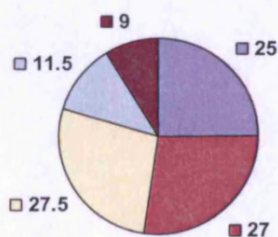
Baseline (29/1/03)

Follow-up (22/8/03)

Levels



Weights



Summary quantitative data

Assessment	Baseline	Follow-up
SEIQoL (total)	42.11	65.83
EQ-5D	Pretest: 0.62	Con. Posttest: 0.88 Thentest: 0.62
Life Orientation Test	17	

Appendix IX Individual case summaries

Patient #23

Age: 46 **Gender:** Female **Status:** Arthroscopy **Job:** Hospital cleaner

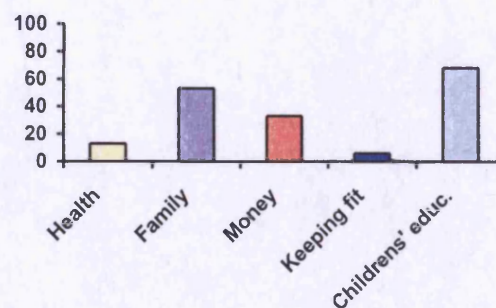
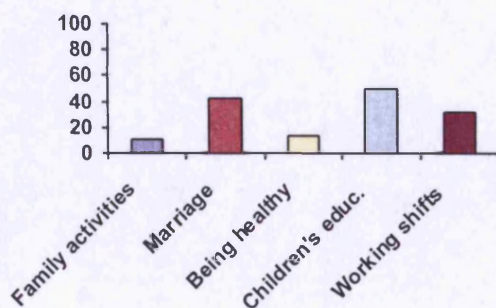
Summary: The patient was listed at baseline for arthroscopy. At baseline she described an approximately 13 year history of problems with her knee, and weakness in the joint. She injured her knee again about 18 months ago whilst playing judo and she reported problems at work with her knee aching. She described having therapy for ligament damage and following a scan she had been referred to orthopaedics by her physiotherapist. Following her arthroscopy she was told that she had no ligament and would require a further operation to reconstruct it.

SEIQoL cues, levels and weights

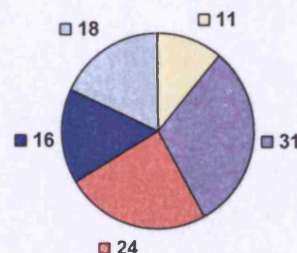
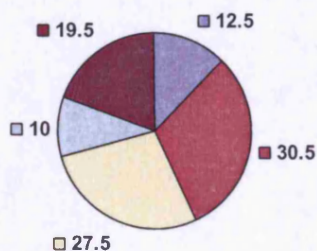
Baseline (30/1/03)

Follow-up (11/8/03)

Levels



Weights



Summary quantitative data

Assessment	Baseline	Follow-up
SEIQoL (total)	29.15	38.98
EQ-5D	Pretest: 0.09	Con. Posttest: 0.73 Thentest: 0.69
Life Orientation Test	9	

Appendix IX Individual case summaries

Patient #24

Age: 56 Gender: Male Status: Arthroscopy Job: Glazier

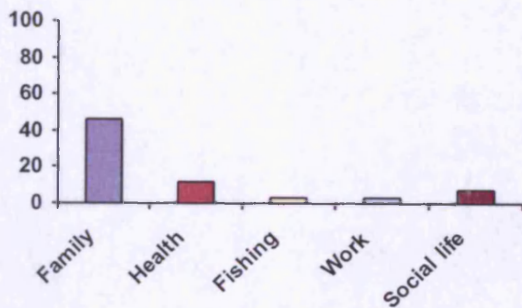
Summary: The patient was listed at baseline for arthroscopy. At baseline he described a three year history with his knee which initially was investigated by x-ray. He reported stiffness, pain, clicking and locking of the joint. Following deteriorating symptoms he had an MR scan and osteoarthritis was diagnosed. He was told that he required a total knee replacement – which he subsequently had. However he has continued to have problems and required an exploratory arthroscopy. At follow-up he reported that his knee still left him in constant pain. He was to have further investigation of his lower back (as a potential source of his knee pain). He is sceptical and resentful of his medical carers and the whole experience of his knee problem and its treatment has left him depressed.

SEIQoL cues, levels and weights

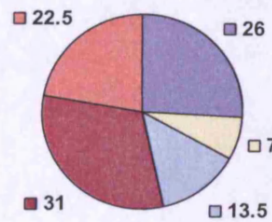
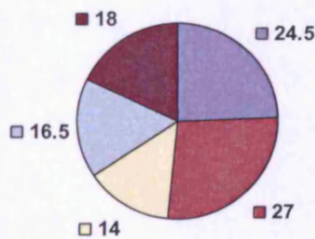
Baseline (31/1/03)

Follow-up (11/8/03)

Levels



Weights



Summary quantitative data

Assessment	Baseline	Follow-up
SEIQoL (total)	16.86	28.29
EQ-5D	Pretest: -0.24	Con. Posttest: -0.07 Thentest: -0.24
Life Orientation Test	12	

Appendix IX Individual case summaries

Patient #25

Age: 35 **Gender:** Male **Status:** Arthroscopy **Job:** Not determined

Summary: At baseline the patient was listed for meniscal repair (ACL reconstruction) and he described a football injury two years previously. He reported clicking and pain in the joint. At the follow-up interview the patient described his operation. Although he appeared to suggest that he had received a total knee replacement, he subsequently described ligament repair. Following the operation, he also had the joint washed out on two occasions, possibly due to an infection. Functionally, he did not appear inhibited and he described going to the gym and conducting his own physiotherapy exercises. He did though comment that he was no longer playing football.

English was not the patient's first language and additionally he may have had a mild speech impediment. He was apparently suspicious of being recorded (although nominally agreeable to it) and it was very difficult to draw the patient into discussion in line with the interview schedule. Neither interview was therefore recorded and formal assessments (SEIQoL, EQ-5D, LOT-R) were not attempted, in favour of a discussion of the history of the knee problem and its treatment.

SEIQoL cues, levels and weights

Baseline (4/2/03)	Follow-up (2/9/03)
-------------------	--------------------

Levels

<i>SEIQoL not completed</i>	<i>SEIQoL not completed</i>
-----------------------------	-----------------------------

Weights

Cue changes:

Summary quantitative data

Assessment	Baseline	Follow-up
SEIQoL (total)	-	-
EQ-5D	<i>Pretest: -</i>	<i>Con. Posttest: - Thentest: -</i>
Life Orientation Test	-	-

Appendix IX Individual case summaries

Patient #26

Age: 51 **Gender:** Female **Status:** Referral **Job:** Not working

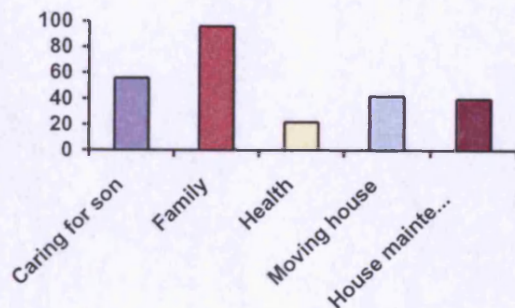
Summary: At baseline the patient had been referred to radiology (MRI) due to pain and giving way in the knee. At the baseline interview she described how she had injured her back whilst gardening the previous year (although she later described how she had injured the ligaments in the same knee when she had been 12). After being 'laid up' for a fortnight she described being unable to straighten her leg. She described how she subsequently had problems with her knee giving way. An x-ray apparently showed arthritis. When asked why the GP referred her for MRI following the x-ray, she mentioned cartilages. She reported stiffness in the joint when she awoke in the morning and pain. Her knee problem was affecting her walking, which she reported doing less of (as well as aerobics and gardening). She considered that the depression, for which she was currently taking medication, had been the result of the knee problem and its impact on her functioning.

SEIQoL cues, levels and weights

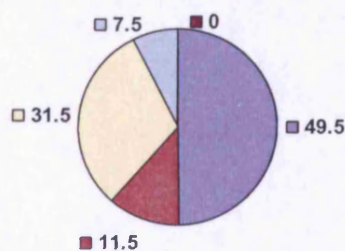
Baseline (5/2/03)

Follow-up (not completed)

Levels



Weights



Summary quantitative data

Assessment	Baseline	Follow-up
SEIQoL (total)	48.84	
EQ-5D	Pretest: 0.19	Con. Posttest: - Thentest: -
Life Orientation Test	-	

Appendix IX Individual case summaries

Patient #27

Age: 37 **Gender:** Female **Status:** Arthroscopy **Job:** Not employed

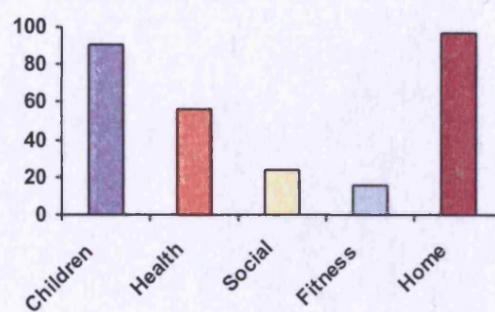
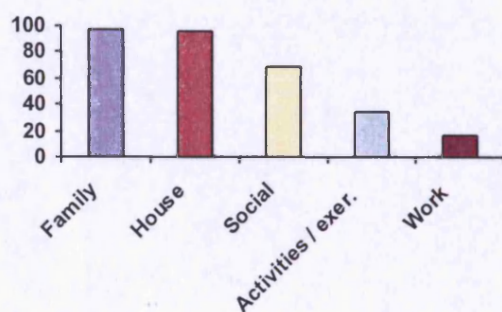
Summary: At baseline the patient was listed for arthroscopy (for removal of tibial screw). The patient described a long history and had undergone ligament surgery the previous year. She reported that the forthcoming operation was to repair the cartilage, and was to complete outstanding from the previous operation. At follow-up (actually 2-3 months following her operation) she reported improvement in her knee. However, she also reported some deterioration of symptoms in her other knee (which had started to become apparent about twelve months previously). These problems are now causing restrictions in her activities, including her general mobility.

SEIQoL cues, levels and weights

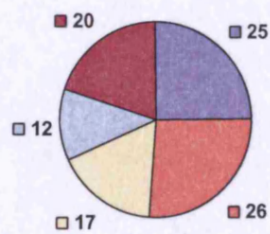
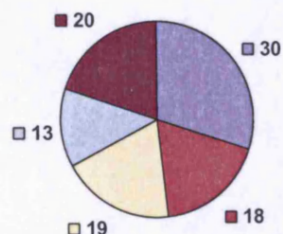
Baseline (5/2/03)

Follow-up (14/8/03)

Levels



Weights



Summary quantitative data

Assessment	Baseline	Follow-up
SEIQoL (total)	67.11	62.71
EQ-5D	Pretest: 0.69	Con. Posttest: 0.62 Thentest: 0.62
Life Orientation Test	14	

Appendix IX Individual case summaries

Patient #28

Age: 61 **Gender:** Female **Status:** Arthroscopy **Job:** Not employed (former nurse)

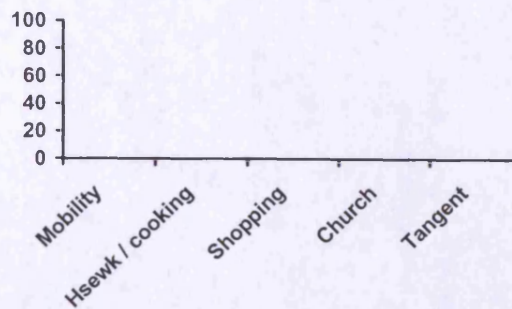
Summary: At baseline the patient was listed for arthroscopy. The patient had substantial co-morbidity (for example, she described rheumatoid and osteo-arthritis) and was physically very frail. She was also taking anti-depressant medication and morphine. She reported that her forthcoming operation was for a partial knee replacement. At follow-up she described how she recently had an operation upon her hand to ease some of her arthritic problems. She also had lost her hair since the baseline interview due to treatment (with Leflunomide). She was now due to have a privately-funded total knee replacement.

SEIQoL cues, levels and weights

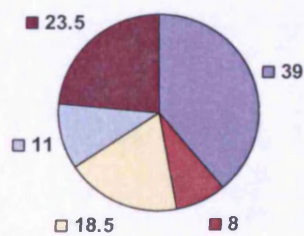
Baseline (5/2/03)

Follow-up (21/8/03)

Levels



Weights



Note: SEIQoL cues only elicited – levels and weights not determined

Summary quantitative data

Assessment	Baseline	Follow-up
SEIQoL (total)	-	-
EQ-5D	Pretest: -0.24	Con. Posttest: -0.19 Thentest: 0.52
Life Orientation Test	13	

Appendix IX Individual case summaries

Patient #29

Age: 54 **Gender:** Female **Status:** Arthroscopy **Job:** Catering worker

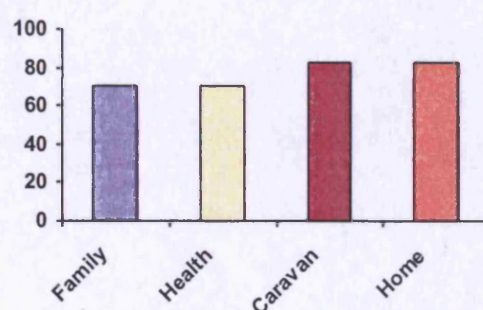
Summary: At baseline the patient had been listed for diagnostic arthroscopy. By the time of the actual interview she had already undergone the procedure and was in the early stages of recovery. She considered that the problem had originally stemmed from an injury at work to her foot. She had been referred by her general practitioner who had suggested that she had cartilage problems. She was having continuing pain in her knee. Her work was also affected. At follow-up she described the operation during which some bone had been removed from behind the cartilage and other debris had been washed out. She reported that her recovery from the operation had been slow, but that in the last 2-3 months that she had been better. She felt that her knee caused her minimal problems functionally.

SEIQoL cues, levels and weights

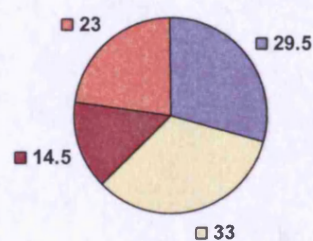
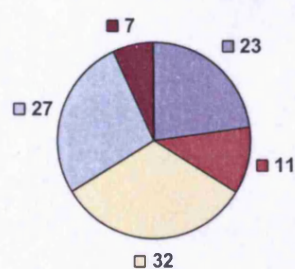
Baseline (7/2/03)

Follow-up (5/9/03)

Levels



Weights



Summary quantitative data

Assessment	Baseline	Follow-up
SEIQoL (total)	87.52	74.88
EQ-5D	Pretest: 0.12	Con. Posttest: 1.00 Thentest: 0.09
Life Orientation Test	15	

Appendix IX Individual case summaries

Patient #30

Age: 44 **Gender:** Female **Status:** Arthroscopy **Job:** Residential care team leader

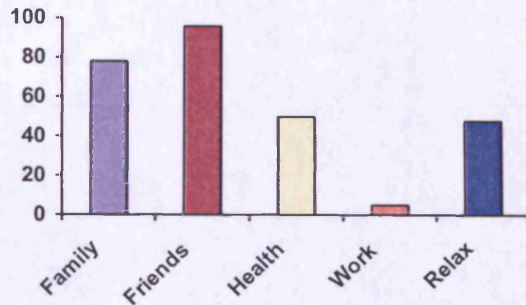
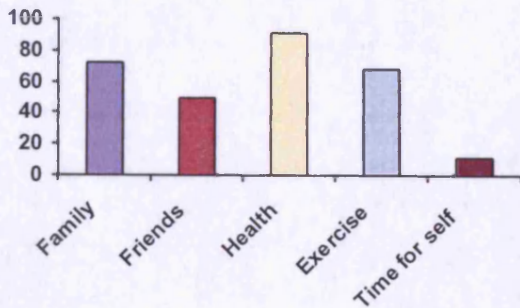
Summary: At baseline the patient was listed for arthroscopy and she reported a diagnosis of cartilage tear. However, she reported no substantial (symptomatic) problem, which she considered was the result of effective weight loss / management. Although given a date for the operation her intention was to defer until her knee became a problem again. At follow-up, she had not had the operation. She had put on weight due to a lack of (motivation to) exercise and increased consumption. She had also been off work for much of the interim due a suspension (subsequent to which she had been fully exonerated) and then stress. She had requested that she be returned to the waiting list as she has now experiencing symptoms following exercise, to which she has recently returned.

SEIQoL cues, levels and weights

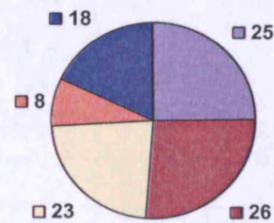
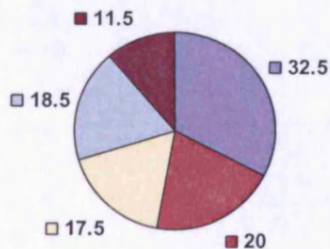
Baseline (11/2/03)

Follow-up (6/11/03)

Levels



Weights



Summary quantitative data

Assessment	Baseline	Follow-up
SEIQoL (total)	63.09	65.00
EQ-5D	Pretest: 0.85	Con. Posttest: 0.85 Thentest: 1.00
Life Orientation Test	10	

Appendix IX Individual case summaries

Patient #31

Age: 57 **Gender:** Male **Status:** Referred **Job:** Newspaper editor

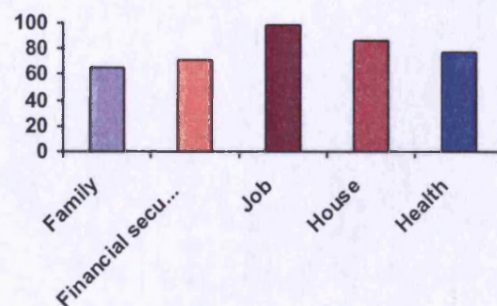
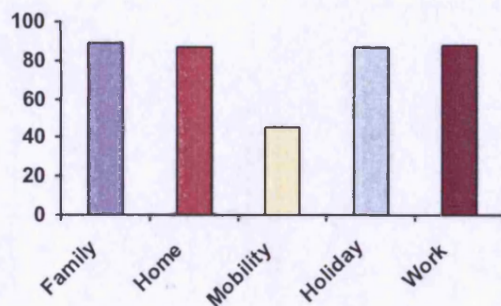
Summary: The referral to radiology at baseline described knee pain, and mentioned the meniscus and an injury. The patient described an active life when he was younger and reported having played professional football. He described having cartilages removed from his right knee some time ago and more recently injuring his left knee. Subsequently, he had been experiencing pain and had developed a limp. Following physiotherapy he was advised he had ligament damage but was referred by his general practitioner for MRI investigation. At follow-up, and following further physiotherapy his limp had gone, although he still had some pain left

SEIQoL cues, levels and weights

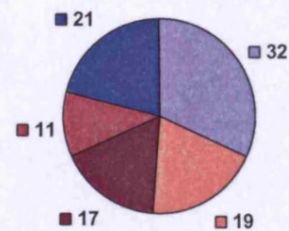
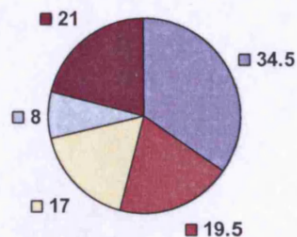
Baseline (13/2/03)

Follow-up (21/8/03)

Levels



Weights



Summary quantitative data

Assessment	Baseline	Follow-up
SEIQoL (total)	80.76	76.79
EQ-5D	Pretest: 0.69	Con. Posttest: 0.73 Thentest: 0.13
Life Orientation Test	14	

Appendix IX Individual case summaries

Patient #32

Age: 46 **Gender:** Male **Status:** Referral **Job:** IT project manager

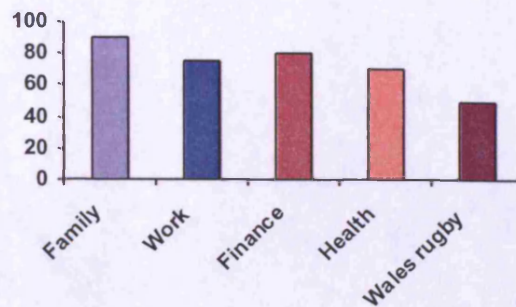
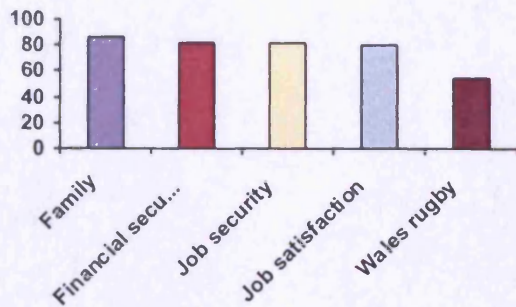
Summary: At baseline the patient had been referred to radiology for MRI. The investigation request had made reference to the meniscus and knee pain. He originally described a gradual onset over the previous two years, with discomfort and some impaired mobility, pain and a sensation of pins and needles. He was concerned especially about his ability to play and care for his young daughter. At follow-up he reported moderate improvement in his knee and that he was less concerned about it. However, he has avoided certain activities which may aggravate the complaint (e.g. home improvements). His main concern was to clarify a diagnosis for his knee.

SEIQoL cues, levels and weights

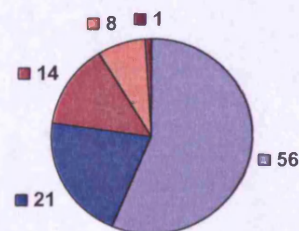
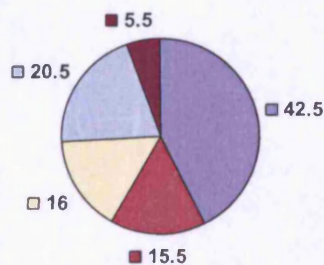
Baseline (14/2/03)

Follow-up (11/9/03)

Levels



Weights



Summary quantitative data

Assessment	Baseline	Follow-up
SEIQoL (total)	82.44	83.45
EQ-5D	Pretest: 1.00	Con. Posttest: 0.88 Thentest: 1.00
Life Orientation Test	20	

Patient #33

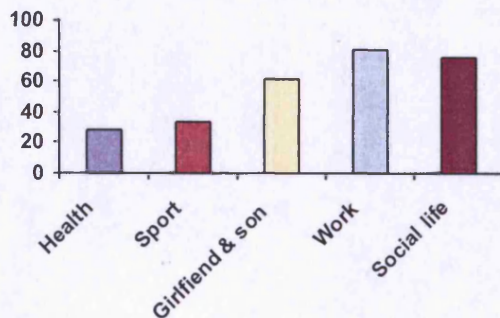
Age: 38 **Gender:** Male **Status:** Referred **Job:** Builder

Summary: At baseline the patient had been referred for radiological investigation with MRI following a rugby injury to his knee. The investigation request made reference to pain, locking and the meniscus. The patient was a builder and keen sportsman, although he now participated in the latter as a referee. He had a long-standing back problem and at the time of the baseline interview he reported some relationship difficulties with his partner which had been exacerbated by a recent miscarriage. He reported that his knee had been assessed in hospital and he had been told that he had ligament damage. He had been prescribed physiotherapy. He was quite critical of the lack of appreciation by clinicians of the impact of his knee problem. He did not respond to repeated requests for a follow-up interview.

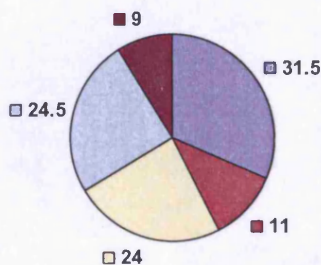
SEIQoL cues, levels and weights

Baseline (14/2/03) **Follow-up (n/a)**

Levels



Weights



Summary quantitative data

Quantitative Assessment	Baseline	Follow-up
SEIQoL (total)	54.02	-
EQ-5D	Pretest: 0.59	Con. Posttest: na Thentest: na
Life Orientation Test	na	

Patient #34

Age: 46 **Gender:** Male **Status** Referred **Job:** Water mains manual worker

Summary: At baseline the patient had been referred to radiology for investigation by MRI. The investigation request referred to pain and the meniscus. He reported a twisting injury to his knee (approximately 13-14 months previously) whilst working and originally having the joint x-rayed. He has been off work ever since the injury. He reported that his walking is affected and that he has constant pain in the joint. As a consequence, he also reported becoming depressed and that he is taking anti-depressant medication. The SEIQoL assessment was not completed at baseline or at follow-up due to the patient's rigid focus upon his knee problem – despite repeated reminders at baseline he continued addressing his knee during its assessment. (However, some 'cues' were discussed at baseline and these are described below.) At follow-up he described how he had changed his job (after being on extended sick-leave, and in receipt of disability benefit) to work which he was more able to perform physically (he also had concomitant physical problems with his neck and back). He reported that his knee was somewhat improved, but that he was taking daily painkillers which have recently been increased in strength.

There were some communication problems – he repeated some points several times (for example, the pain he had and its impact upon him). He reported the adverse effects of his physical problems upon his ability to concentrate (and his mental health in general) and during the interview he often took a long time to answer questions.

SEIQoL cues, levels and weights

Baseline (18/2/03)	Follow-up (12/11/03)
Levels	
n/a	n/a
Weights	
n/a	n/a

Summary quantitative data

Quantitative Assessment	Baseline	Follow-up
SEIQoL (total)	na	na
EQ-5D	<i>Pretest:</i> na	<i>Con. Posttest:</i> 0.19 <i>Thentest:</i> 0.02
Life Orientation Test	11	

Patient #35

Age: 63 **Gender:** Female **Status:** Referred **Job:** Retired social worker

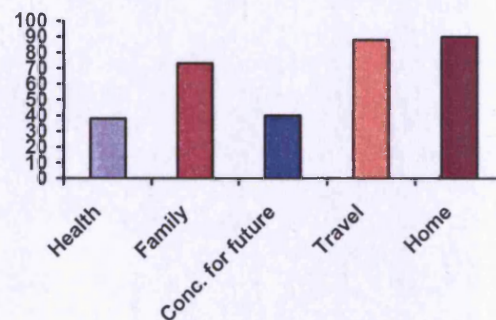
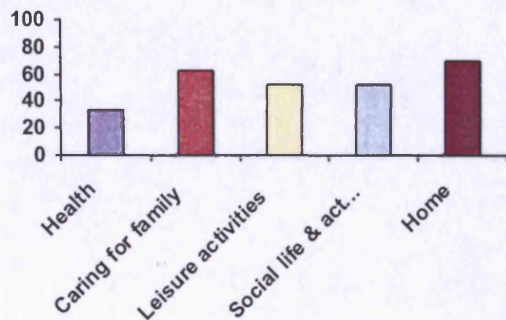
Summary: At baseline the patient had been referred for radiological investigation by MRI. The investigation request form referred to the meniscus, pain, locking and giving way. In conversation she confirmed that her knee had been giving way and been painful recently. She also described a long history of arthritis in both knees. She was concerned about the broad impact on her life that the knee problem was causing her. She had a number of concomitant medical conditions including diabetes, duodenal ulcers and recurrent shingles. At follow-up she reported that her knee was somewhat worse, and that it was giving way a bit more and was more painful. She reported that she had made some adaptations in her behaviour to cope with the impact upon her mobility. At follow-up she was still waiting for her MRI.

SEIQoL cues, levels and weights

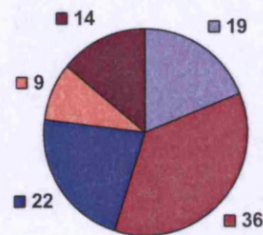
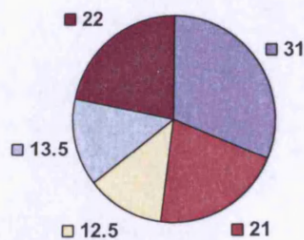
Baseline (20/02/03)

Follow-up (3/11/03)

Levels



Weights



Summary quantitative data

Assessment	Baseline	Follow-up
SEIQoL (total)	52.38	62.82
EQ-5D	Pretest: 0.62	Con. Posttest: 0.62 Thentest: 0.62
Life Orientation Test	22	

Appendix IX Individual case summaries

Patient #36

Age: 40 **Gender:** Male **Status:** Arthroscopy **Job:** Builder

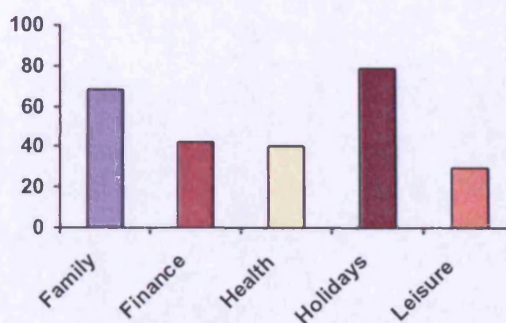
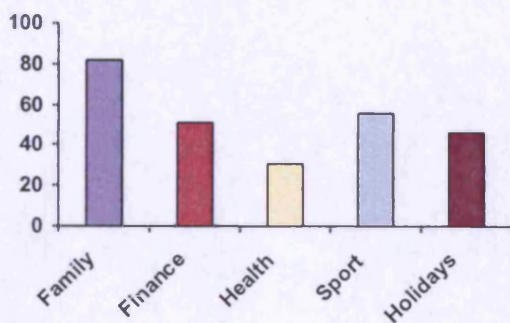
Summary: At baseline the patient was listed for arthroscopy, and at the time described a long history of knee problems following accidents and sporting activities when younger. He also described a lifting injury in work three years ago, following which he was temporarily unable to work and subsequently suffered swelling and pain. His general practitioner suspected a torn cartilage, which was subsequently confirmed by MRI. At baseline he had already had the operation on his knee. At follow-up he reported that he had been told he would require further surgery, which he had temporarily postponed. This was because he was concerned about being incapacitated for a substantial period of time; being unable to work; and the impact of their finances. He reported that the knee was now less swollen, but also less mobile and more painful. The patient, a builder, was interviewed on both occasions with his wife present. She often volunteered information and even played a role in probing and occasionally contradicting the patient during the interview.

SEIQoL cues, levels and weights

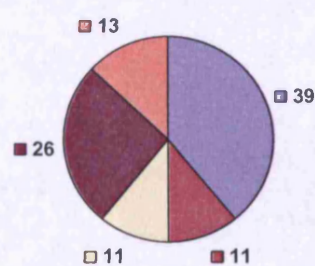
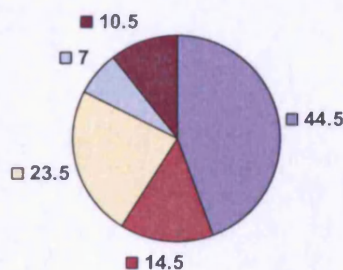
Baseline (20/2/03)

Follow-up (7/11/03)

Levels



Weights



Summary quantitative data

Assessment	Baseline	Follow-up
SEIQoL (total)	59.92	59.98
EQ-5D	Pretest: 0.62	Con. Posttest: 0.69 Thentest: 0.69
Life Orientation Test	10	

Appendix IX Individual case summaries

Patient #37

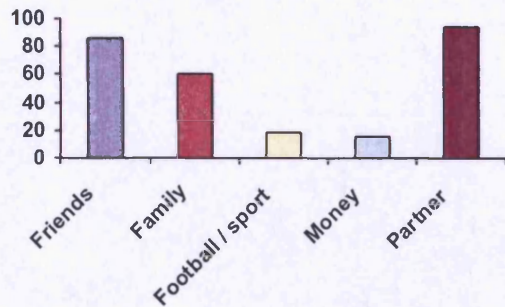
Age: 26 **Gender:** Male **Status** Referred **Job:** Insurance company administrator

Summary: At baseline the patient had been referred to radiology for MRI of the knee. The investigation form referred to the meniscus and pain in the knee. The baseline interview was not recorded and the patient could not be followed-up for second interview. Therefore, assessment of cue change was not possible.

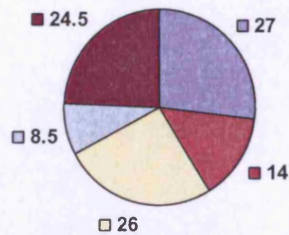
SEIQoL cues, levels and weights

Baseline (26/3/03) **Follow-up (not completed)**

Levels



Weights



Summary quantitative data

Assessment	Baseline	Follow-up
SEIQoL (total)	60.95	
EQ-5D	Pretest: 0.69	Con. Posttest: - Thentest: -
Life Orientation Test	-	

Appendix IX Individual case summaries

Patient #38

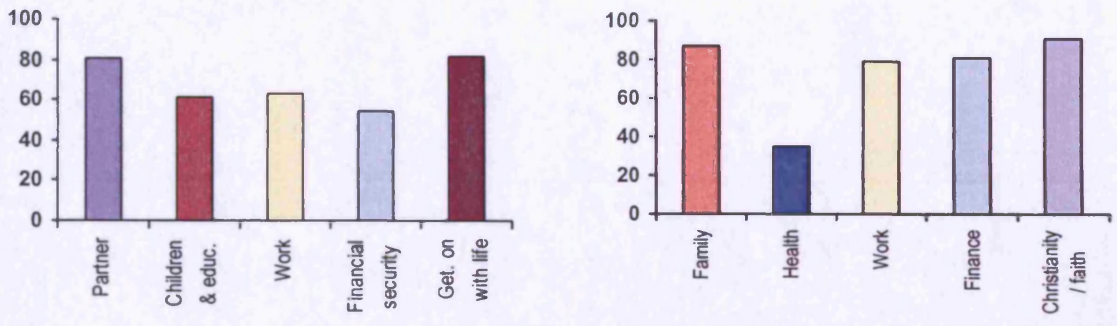
Age: 47 **Gender:** Female **Status:** Referral **Job:** Teacher

Summary: At baseline the patient had been referred to radiology for MRI of the knee. She described an old injury to the knee (18 years previously) but with no intervening problems. More recently she had injured her knee, falling on it at home a few months previously. Initially she limped for about a fortnight afterwards. She also had stiffness in the joint; it had given way, and walking and driving were painful. Her general practitioner requested MRI but did not offer a diagnosis. At follow-up she reported her knee was worse (and now included locking) and that her other knee had also started causing her problems. She had started avoiding certain activities which she knew would be problematic otherwise.

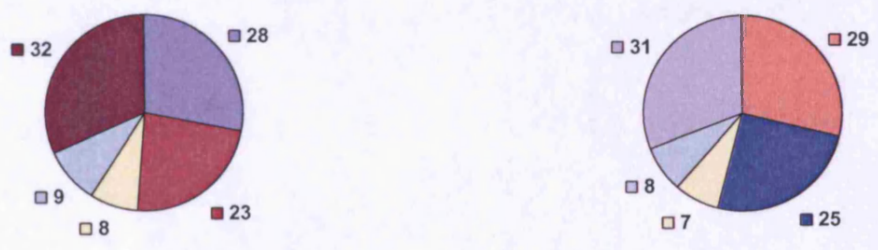
SEIQoL cues, levels and weights

Baseline (1/4/03) Follow-up (4/11/03)

Levels



Weights



Summary quantitative data

Assessment	Baseline	Follow-up
SEIQoL (total)	72.85	74.20
EQ-5D	Pretest: 0.69	Con. Posttest: 0.69 Thentest: 0.81
Life Orientation Test	21	

Appendix IX Individual case summaries

Patient #39

Age: 50 **Gender:** Female **Status:** Referral **Job:** Local government administrator

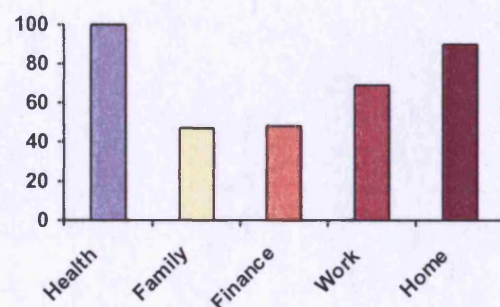
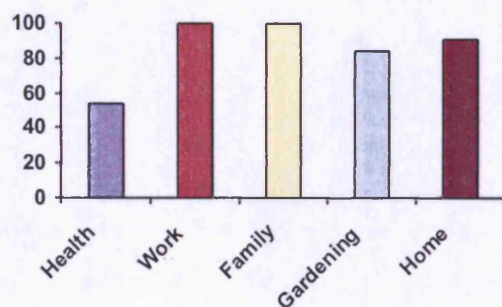
Summary: At baseline patient had been referred for radiological investigation with MRI. Her request form referred to the meniscus, knee pain but her general practitioner had also speculated about a Baker's cyst. She reported an onset of problems six months previously, including pain and locking, but following no obvious injury. Subsequently, she had experienced monthly episodes which persist for about a week. At follow-up, she reported that her knee had improved and was less problematic than it had been previously. Both knees ached, and she considered that she had a Baker's cyst in each, and cartilage damage in her (most) affected knee.

SEIQoL cues, levels and weights

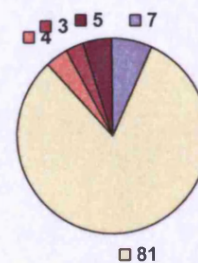
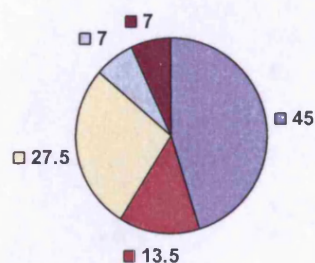
Baseline (2/4/03)

Follow-up (30/10/03)

Levels



Weights



Summary quantitative data

Assessment	Baseline	Follow-up
SEIQoL (total)	77.5	53.56
EQ-5D	Pretest: 0.59	Con. Posttest: 0.81 Thentest: 0.62
Life Orientation Test	13	

Published letters

Robling MR, Hood K, Butler C. Interpreting differences over time in patient self-evaluation. *British Journal of General Practice* 2005: **55** (515); 472

Robling MR, Hood K. Response shift, responsiveness or recall bias? *British Journal of General Practice* 2002: **52** (480); 585

Conference presentations (Oral):

Robling MR. Compounding the problem, confounding the outcome? Waiting for knee care in South Wales. Society for Academic Primary Care Annual Scientific Meeting (South West Region), 2005, Cardiff

Robling MR. The impact of knee problems upon quality of life: a case study of patient response shift. Society for Academic Primary Care Annual Scientific Meeting (South West Region), 2004, Bristol

Robling MR. The impact of internal derangement of the knee upon patient quality of life: in context and over time. Wales Primary Care Symposium, 2004 Llandrindod Wells

Robling MR. The impact of knee problems upon quality of life: change and accommodation within the patient biography. British Sociological Association: Medical Sociology Annual Scientific Meeting, 2003, York.

Robling MR. Measuring changes in quality of life over time: the role of response shift. 17th Annual Postgraduate Research Day, University of Wales College of Medicine, 2002, Cardiff.

Conference presentations (Poster):

Robling M. Spot the difference: assessing quality of life scale responsiveness – an example from general practice research. 15th Annual Postgraduate Research Day, University of Wales College of Medicine 2000, Cardiff.

Robling M. Spot the difference: assessing quality of life scale responsiveness – an example from general practice research. Association of University Departments of General Practice Annual Scientific Meeting, 2000, Bournemouth.

Bibliography

1. The BRIDGE Study Group. The presentation and management of breast symptoms in general practice in South Wales - The BRIDGE Study. *British Journal of General Practice* 1999;49:811-812.
2. The BRIDGE Study Group. Responses of primary health care professionals to the UK national guidelines on management and referral of women with breast conditions. *Journal of Evaluation in Clinical Practice* 2002;8(3):319-325.
3. Ware J, Sherbourne C. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical Care* 1992;30(6):473-83.
4. The EuroQol Group. EuroQol - a new facility for the measurement of health-related quality of life. *Health Policy* 1990;16:199-208.
5. Robling M, Matthews S, Hood K, Russell I, Holloway R, Wilkinson C, et al. The development of a new site-specific measure of quality of life for breast problems: The Cardiff breast scales. *Quality of Life Research* 2002;11:339-348.
6. Streiner DL, Norman GR. *Health measurement scales: a practical guide to their development and use*. Second ed. Oxford: Oxford University Press, 2000.
7. Fitzpatrick R, Davey C, Buxton MJ, Jones D. Evaluating patient-based outcome measures for use in clinical trials. *Health Technology Assessment* 1998;2(4).
8. Jobe JB. Cognitive psychology and self-reports: models and methods. *Quality of Life Research* 2003;12:219-227.
9. Schwartz CE, Sprangers MAG. *Adaptation to Changing Health: Response Shift in Quality-of-life Research*. Washington: American Psychological Association, 2000.
10. Sprangers MAG, Schwartz CE. Integrating response shift into health-related quality of life research: a theoretical model. *Social Science and Medicine* 1999;48:1507-1515.
11. Schwartz CE, Sprangers MAG. Methodological approaches for assessing response shift in longitudinal health-related quality of life research. *Social Science and Medicine* 1999;48:1531-1548.
12. Terwee CB, Dekker FW, Wiersinga W, Prummel M, Bossuy P. On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. *Quality of Life Research* 2003;12:349-362.
13. Kirshner B, Guyatt G. A methodological framework for assessing health indices. *Journal of Chronic Diseases* 1985;38(1):27-36.
14. Testa MA, Simonson DC. Assessment of quality-of-life outcomes. *New England Journal of Medicine* 1996;334(13):835--840.
15. Aaronson NK. Quality of life assessment in clinical trials. *Controlled Clinical Trials* 1989;10:195S-208S.
16. Hays RD, Anderson R, Revicki D. Psychometric considerations in evaluating health-related quality of life measures. *Quality of Life Research*, 1993:441-449.
17. Deyo RA. Measuring functional outcomes in therapeutic trials for chronic diseases. *Controlled Clinical Trials* 1984;5:223-240.

18. Beaton DE, Bombardier C, Katz JN, Wright JG. A taxonomy for responsiveness. *Journal of Clinical Epidemiology* 2001;54:1204-1217.
19. de Bruin AF, Diederiks JPM, de Witte LP, Stevens FCJ, H P. Assessing the responsiveness of a functional status measure: the Sickness Impact Profile versus the SIP68. *Journal of Clinical Epidemiology* 1997;50(5):529-540.
20. Katz JN, Larson MG, Phillips CB, Fossel AH, Liang MH. Comparative measurement sensitivity of short and longer health status instruments. *Medical Care* 1992;30(10):917- 925.
21. Husted JA, Cook RJ, Farewell V, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. *Journal of Clinical Epidemiology* 2000;53:459-468.
22. Stratford PW, Binkley JM, Riddle DL. Health status measures: strategies and analytic methods for assessing change scores. *Physical Therapy* 1996;76(10):1109-1123.
23. Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. *Journal of Chronic Diseases* 1986;39(11):897-906.
24. Lydick E, Epstein RS. Interpretation of quality of life changes. *Quality of Life Research* 1993;2:221-226.
25. Liang MH. Evaluating measurement responsiveness. *The Journal of Rheumatology* 1995;22:1191-1192.
26. Fortin P, Stucki G, Katz J. Measuring relevant change: an emerging challenge in rheumatologic clinical trials. *Arthritis and Rheumatism* 1995;38(8):1027-1030.
27. Norman G, Stratford P, Regehr G. Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. *Journal of Clinical Epidemiology* 1997;50(8):869-879.
28. Murawski MM, Miederhoff PA. On the generalizability of statistical expressions of health related quality of life instrument responsiveness: a data synthesis. *Quality of Life Research* 1998;7:11-22.
29. O'Keeffe S, Lye M, Donnellan, DN C. Reproducibility and responsiveness of quality of life assessment and six minute walk test in elderly heart failure patients. *Heart* 1998;80(377):382.
30. Beaton DE, Hogg-Johnson S, Bombardier C. Evaluating changes in health status: reliability and responsiveness of five generic health status measures in workers with musculoskeletal disorders. *Journal of Clinical Epidemiology* 1997;50(1):79-93.
31. Hays RD, Hadorn D. Responsiveness to change: an aspect of validity, not a separate dimension. *Quality of Life Research* 1992;1:73-75.
32. Hays RD, Revicki D. Reliability and validity (including responsiveness). In: Fayers P, Hays R, editors. *Assessing quality of life in clinical trials*. Second ed. Oxford: Oxford University Press, 2005.
33. Kim S, Hays RD, Birbeck GL, G VB. Responsiveness of the Quality of Life in Epilepsy Inventory (QOLIE-89) in an Antiepileptic Drug Trial. *Quality of Life Research* 2003;12:147-155.
34. Guyatt GH, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *Journal of Chronic Diseases* 1987;40(2):171-178.

35. Guyatt G, Deyo RA, Charlson M, N LM, Mitchell A. Responsiveness and validity in health status measurement: a clarification. *Journal of Clinical Epidemiology* 1989;42(5):403-408.
36. Corzillius M, Fortin P, Stucki G. Responsiveness and sensitivity to change of SLE disease activity measures. *Lupus* 1999;8:655-659.
37. Guyatt G, Kirshner B, Jaeschke R. Measuring health status: what are the necessary measurement properties? *Journal of Clinical Epidemiology* 1992;45(12):1341-1345.
38. Nunnally JC, Bernstein IH. *Psychometric Theory*. Third Edition ed. New York: McGraw-Hill Inc, 1994.
39. Kazis LE, Anderson JJ, Meenan R. Effect sizes for interpreting changes in health status. *Medical Care* 1989;27(3 (Supplement)):S178-S189.
40. Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopaedic evaluation. *Medical Care* 1990;28(7):632-642.
41. Liang MH. Longitudinal construct validity. *Medical Care* 2000;38(9 (Supplement II)):II-84-II-90.
42. Liang MH, Larson MG, Cullen KE, Schwartz J. Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. *Arthritis and Rheumatism* 1985;28(5):542-547.
43. Wiebe S, Guyatt G, Weaver B, Matijevic S, Sidwell C. Comparative responsiveness of generic and specific quality-of-life instruments. *Journal of Clinical Epidemiology* 2003;56:52-60.
44. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Second Edition ed. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1988.
45. Hevey D, McGee H. The effect size statistic: useful in health outcomes research? *Journal of Health Psychology* 1998;3(2):163-170.
46. Lipsey M. A scheme for assessing measurement sensitivity in program evaluation and other applied research. *Psychological Bulletin* 1983;94(1):152-165.
47. Garratt AM, Moffett JK, Farrin A. Responsiveness of generic and specific measures of health outcome in low back pain. *Spine* 2001;26(1):71-77.
48. Bessette L, Sangha O, Kuntz K, Keller RB, Lew RA, Fossel AH, et al. Comparative responsiveness of generic versus disease-specific and weighted versus unweighted health status measures in carpal tunnel syndrome. *Medical Care* 1998;36(4):491-502.
49. Stucki G, Liang MH, Fossel A, Katz JN. Relative responsiveness of condition-specific and generic health status measures in degenerative lumbar spinal stenosis. *Journal of Clinical Epidemiology* 1995;48(11):1369-1378.
50. Wright JG, Young NL. A comparison of different indices of responsiveness. *Journal of Clinical Epidemiology* 1997;50(3):239-246.
51. Feise RJ, Menke JM. Functional rating Index: a new valid and reliable instrument to measure the magnitude of clinical change in spinal conditions. *Spine* 2001;26(1):78-87.
52. Guyatt GH, Townsend M, Berman LB, Keller JL. A comparison of likert and visual analogue scales for measuring change in function. *Journal of Chronic Diseases* 1987;40(12):1129-1133.

53. Nagata C, Ido M, Shimizu H, Misao A, Matsuura H. Choice of response scale for health measurement: comparison of 4, 5, and 7-point scales and visual analogue scale. *Journal of Epidemiology* 1996;6(4):192-197.
54. Rouf J, Sangha O, Stucki G. Comparative responsiveness of 3 functional indices in ankylosing spondylitis. *The Journal of Rheumatology* 1999;26(9):1959-1963.
55. Vermeersch DA, Lambert MJ, Burlingame GM. Outcome questionnaire: Item sensitivity to change. *Journal of Personality Research* 2000;74(2):242-261.
56. Von Korff M, Ormel J, Keefe FJ, Dworkin SF. Grading the severity of chronic pain. *Pain* 1992;50:133-149.
57. Robling MR, Hood K. Response shift, responsiveness or recall bias? *British Journal of General Practice* 2002;52(480):585.
58. Smith H, Taylor R, Mitchell A. A comparison of four quality of life instruments in cardiac patients: SF-36, QLI, QLMI, and SEIQoL. *Heart* 2000;84:390-394.
59. Moons P, Marquet K, Budts W, de Geest S. Validity, reliability and responsiveness of the "Schedule for the Evaluation of Individual Quality of Life - Direct Weighting" (SEIQoL-DW) in congenital heart disease. *Health and Quality of Life Outcomes* 2004;2:27.
60. Kern RZ, Brown AD. Disease adaptation may have decreased quality-of-life responsiveness in patients with chronic progressive neurological disorders. *Journal of Clinical Epidemiology* 2004;57:1033-1039.
61. Moran LA, Guyatt GH, Norman GR. Establishing the minimal number of items for a responsive, valid, health-related quality of life instrument. *Journal of Clinical Epidemiology* 2001;54(6):571-579.
62. van der Heijden GJMG, Leffers P, Boute LM. Shoulder disability questionnaire design and responsiveness of a functional status measure. *Journal of Clinical Epidemiology* 2000;53:29-38.
63. Montgomery SA, Åsberg M. A new depression scale designed to be sensitive to change. *British Journal of Psychiatry* 1979;134:382-389.
64. Chambers LW, Haight M, Norman G, MacDonald L. Sensitivity to change and the effect of mode of administration on health status measurement. *Medical Care* 1987;25(6):470-480.
65. Schwartz CE, Rapkin BD. Reconsidering the psychometrics of quality of life assessment in light of response shift and appraisal. *Health and Quality of Life Outcomes* 2004;2(16).
66. Rockwood K, Howlett S, Stadnyk K, Carver D, Powell C, Stolee P. Responsiveness of goal attainment scaling in a randomized controlled trial of comprehensive geriatric assessment. *Journal of Clinical Epidemiology* 2003;56:736-743.
67. Blanchard C, Feeny D, Mahon JL, Bourne R, Rorabeck C, Stitt L, et al. Is the Health Utilities Index responsive in total hip arthroplasty patients? *Journal of Clinical Epidemiology* 2003;56:1046-1054.
68. Ruperto N, Ravelli A, Migliavacca D, Viola S, Pistorio A, Duarte C, et al. Responsiveness of clinical measures in children with oligoarticular juvenile chronic arthritis. *The Journal of Rheumatology* 1999;26:1827-1830.

69. Cohen J. *Statistical power analysis for the behavioural sciences*. Revised ed. Orlando: Academic Press Inc., 1977.
70. Middel B, von Sonderen E. Statistical significant change versus relevant or important change in (quasi) experimental design: some conceptual and methodological problems in estimating magnitude of intervention-related change in health services research. *International Journal of Integrated Care* 2002;2.
71. Anderson JJ, Chernoff MC. Sensitivity to change of rheumatoid arthritis clinical trial outcome measures. *Journal of Rheumatology* 1993;20:535-537.
72. Jones G, Jenkinson C, Kennedy S. Evaluating the responsiveness of the endometriosis health profile questionnaire. *Quality of Life Research* 2004;13:705-713.
73. Schmitt JS, Di Fabio RP. Reliable change and minimum important difference (MID) proportions facilitated group responsiveness comparisons using individual threshold criteria. *Journal of Clinical Epidemiology* 2004;57:1008-1018.
74. Brazier J, Harper R, Munro J, Walters S, Snaith M. Generic and condition specific outcome measures for people with osteoarthritis of the knee. *Rheumatology* 1999;38:870-877.
75. Marx RG, Jones EC, Allen AA, Altchek DW, O'Brien SJ, Rodeo SA, et al. Reliability, validity, and responsiveness of four knee outcome scales for athletic patients. *Journal of Bone and Joint Surgery* 2001;83-A(10):1459-1469.
76. Barber-Westin SD, Noyes FR, McCloskey JW. Rigorous statistical reliability, validity, and responsiveness testing of the Cincinnati Knee Rating System in 350 subjects with uninjured, injured, or a anterior cruciate ligament-reconstructed knees. *The American Journal of Sports Medicine* 1999;27(4):402-416.
77. Garratt AM, Hutchinson A, Russell I, (NEBPINY) icwtNfE-BPiNaY. The UK version of the Seattle Angina Questionnaire (SAQ-UK): reliability, validity and responsiveness. *Journal of Clinical Epidemiology* 2001;54:907-915.
78. Fitzpatrick R, Ziebland S, Jenkinson C, Mowat A, Mowat A. A comparison of the sensitivity to change of several health status instruments in rheumatoid arthritis. *The Journal of Rheumatology* 1993;20(3):429-436.
79. Walker DJ, Kidd E, Heslop PS, Chandler C. Spontaneous ambulatory activity as a quantifiable outcome measure for rheumatoid arthritis. *Rheumatology* 1999;38:1234-1238.
80. Rose MS, Koshman ML, Spreng S, Sheldon R. Statistical issues encountered in the comparison of health-related quality of life in diseased patients to published general population norms: problems and solutions. *Journal of Clinical Epidemiology* 1999;52(5):405-412.
81. Taylor SJ, Taylor AE, Foy MA, Fogg AJB. Responsiveness of common outcome measures for patients with low back pain. *Spine* 1999;24(17):1805-1812.
82. Salford Software. *FTN95 Users Guide*: Salford Software Ltd, 2000.
83. Norman GR, Sridhar FG, Guyatt GH, Walter SD. Relation of distribution- and anchor-based approaches in interpretation of changes in health-related quality of life. *Medical Care* 2001;39(10):1039-1047.

84. Lindgren BW. *Statistical Theory*. Third ed. New York: Macmillan Publishing Co., Inc., 1976.
85. Guyatt GH, Norman GR, Juniper EF, Griffith LE. A critical look at transition items. *Journal of Clinical Epidemiology* 2002;55:900-908.
86. Cella D, Hahn EA, Dineen K. Meaningful change in cancer-specific quality of life scores: differences between improvement and worsening. *Quality of Life Research* 2002;11:207-221.
87. Lenert LA, Treadwell JR, Schwartz CE. Associations between health status and utilities: indirect evidence for response shift. In: Schwartz CE, Sprangers MA, editors. *Adaptation to changing health: response shift in quality-of-life research*. Washington, DC: American Psychological Association, 2000.
88. Propert KJ, Litwin MS, Wang Y, Alexander RB, Calhoun E, Nickel JC, et al. Responsiveness of the National Institutes of Health Chronic Prostatitis Symptom Index (NIH-CPSI). *Quality of Life Research* 2006;15:299-305.
89. Golembiewski RT, Billingsley K, Yeager S. Measuring change and persistence in human affairs: types of change generated by OD designs. *Journal of Applied Behavioral Science* 1976;12:133-157.
90. Howard GS, Ralph KM, Gulanick NA, Maxwell SE, Nance DW, Gerber SK. Internal invalidity in pretest-posttest self-report evaluations and a re-evaluation of retrospective pretests. *Applied Psychological Measurement* 1979;3(1):1-23.
91. Howard GS, Dailey PR. Response-shift bias: a source of contamination of self-report measures. *Journal of Applied Psychology* 1979;64(2):144-150.
92. Campbell DT, Stanley JC. *Experimental and quasi-experimental designs for research*. Boston: Houghton Mifflin Company, 1966.
93. Schwartz C, Sprangers M, Fayers P. Response shift: you know it's there but how do you capture it? Challenges for the next phase of research. In: Fayers P, Hays R, editors. *Assessing quality of life in clinical trials*. Second ed. Oxford: Oxford Medical Publications, 2005.
94. Schwartz CE, Sprangers MA, Carey A, Reed G. Exploring response shift in longitudinal data. *Psychology and Health* 2004;19(1):51-69.
95. Rapkin BD, Schwartz CE. Toward a theoretical model of quality-of-life appraisal: implications of findings from studies of response shift. *Health and Quality of Life Outcomes* 2004;2(14).
96. Schwartz CE, Sprangers MA. Discussion: theoretical reflections. In: Schwartz CE, Sprangers MA, editors. *Adaptation to changing health: response shift in quality-of-life research*. Washington, DC: American Psychological Association, 2000.
97. Tourangeau R, Rips LJ, Rasinski K. *The Psychology of Survey Response*. Cambridge: Cambridge University Press, 2000.
98. Skeff KM, Stratos GA, Bergen MR. Evaluation of a medical faculty development program: a comparison of traditional pre/post and retrospective pre/post self-assessment ratings. *Evaluation & the health professions* 1992;15(3):350-366.
99. Sprangers M, Hoogstraten J. Response-style effects, response-shift bias and a bogus-pipeline. *Psychological Reports* 1987;61:579-585.

100. Robinson EAR, Doueck HJ. Implications of the pre/post/then design for evaluating social group work. *Research on Social Work Practice* 1994;4(2):224-239.
101. Gibbons F. Social comparison as a mediator of response shift. *Social Science and Medicine* 1999;48:1517-1530.
102. Van der Zee KI, Buunk BP, De Ruiter JH, Tempelaar R, Van Sonderen E, Sanderman R. Social comparison and the subjective well-being of cancer patients. *Basic and Applied Social Psychology* 1996;18(4):453-468.
103. Schwartz CE. Teaching coping skills enhances quality of life more than peer support: results of a randomized trial with multiple sclerosis patients. *Health Psychology* 1999;18(3):211-220.
104. King KB, Clark PC, Friedman MM. Social comparisons and temporal comparisons after coronary artery surgery. *Heart and Lung* 1999;28(5):316-325.
105. Beaton DE, Tarasuk V, Katz JN, Wright JG, Bombardier C. "Are you better?" A qualitative study of the meaning of recovery. *Arthritis Care and Research* 2001;45:270-279.
106. Lepore SJ, Eton DT. Response shift in prostate cancer: an evaluation of suppressor and buffer models. In: Schwartz CE, Sprangers MA, editors. *Adaptation to changing health: response shift in quality-of-life research*. Washington, DC: American Psychological Association, 2000.
107. Hagedoorn M, Sneeuw KCA, Aaronson NK. Changes in physical functioning and quality of life in patients with cancer: Response shift and relative evaluation of one's condition. *Journal of Clinical Epidemiology* 2002;55:176-183.
108. Carver CS, Scheier MF. Scaling back goals and recalibration of the affect system are processes in normal adaptive self-regulation: understanding 'response shift' phenomena. *Social Science and Medicine* 2000;50:1715-1722.
109. Richards TA, Folkman S. Response shift: a coping perspective. In: Schwartz CE, Sprangers MA, editors. *Adaptation to changing health: response shift in quality-of-life research*. Washington, DC: American Psychological Association, 2000.
110. Norman G. Hi! How are you? Response shift, implicit theories and differing epistemologies. *Quality of Life Research* 2003;12:239-249.
111. Brossart DF, Clay DL, Willson VL. Methodological and statistical considerations for threats to internal validity in pediatric outcome data: response shift in self-report outcomes. *Journal of Pediatric Psychology* 2002;27(1):97-107.
112. Schwartz CE, Sendor RM. Helping others helps oneself: response shift effects in peer support. *Social Science and Medicine* 1999;48:1563-1575.
113. Millsap RE, Hartog SB. Alpha, beta, and gamma change in evaluation research: a structural equation approach. *Journal of Applied Psychology* 1988;73(3):574-584.
114. Schwartz C, Sprangers M. Introduction to symposium on the challenge of response shift in social science and medicine. *Social Science and Medicine* 1999;48:1505-1506.

115. Ahmed S, Mayo NE, Wood-Dauphinee S, Hanley JA, Cohen SR. Response shift influenced estimates of change in health-related quality of life poststroke. *Journal of Clinical Epidemiology* 2004;57:561-570.
116. Bernhard J, Lowy A, Mathys N, Herrmann R, Hürny C. Health related quality of life: a changing construct? *Quality of Life Research* 2004;13:1187-1197.
117. Chin K, Fukuhara S, Takahashi K, Sumi K, Nakamura T, Matsumoto H, et al. Response shift in perception of sleepiness in obstructive sleep apnea-hypopnea syndrome before and after treatment with nasal CPAP. *Sleep* 2004;27(3):490-493.
118. Donovan KA, Jacobsen PB, Andrykowski MA, Winters EM, Balducci L, Malik U, et al. Course of fatigue in women receiving chemotherapy and/or radiotherapy for early stage breast cancer. *Journal of Pain and Symptom Management* 2004;28(4):373-380.
119. Holzner B, Kemmler G, Cella D, De Paoli C, Meraner V, Kopp M, et al. Normative data for functional assessment of cancer therapy. *Acta Oncologica* 2004;43(2):153-160.
120. Kidd P, Parshall MB, Wojcik S, Struttmann T. Assessing recalibration as a response shift phenomenon. *Nursing Research* 2004;53(2):130-135.
121. Paterson C. Seeking the patient's perspective: a qualitative assessment of EuroQol, COOP-WONCA charts and MYMOP. *Quality of Life Research* 2004;13:871-881.
122. Roos EJ, de Graeff A, van Eijkeren MA, Boon TA, Heintz APM. Quality of life after pelvic exenteration. *Gynecologic Oncology* 2004;93:610-614.
123. Schwartz CE, Merriman MP, Reed GW, Hammes BJ. Measuring patient treatment preferences in end-of-life care research: applications for advance care planning interventions and response shift research. *Journal of Palliative Medicine* 2004;7(2):233-245.
124. Essink-Bot M, Korfage I, de Koning H. Including the quality-of-life effects in the evaluation of prostate cancer screening: expert opinions revisited? *BJU International* 2003;92(Supplement 2):101-105.
125. Rees J, Waldron D, O'Boyle C, Ewings P, MacDonagh R. Prospective vs retrospective assessment of lower urinary tract symptoms in patients with advanced prostate cancer: the effect of 'response shift'. *BJU International* 2003;92:703-706.
126. Schwartz C, Meisenhelder JB, Ma Y, Reed G. Altruistic social interest behaviors are associated with better mental health. *Psychosomatic Medicine* 2003;65:778-785.
127. Timmerman AA, Anteunis LJ, Meesters CM. Response-shift bias and parent-reported quality of life in children with otitis media. *Archives of Otolaryngology - Head & Neck Surgery* 2003;129:987-991.
128. Greiffenstein MF, Baker WJ. Actual versus self-reported scholastic achievement of litigating postconcussion and severe head injury claimants. *Psychological Assessment* 2002;14(2):202-208.
129. Joore M, Potjewijd J, Timmerman A, Anteunis L. Response shift in the measurement of quality of life in hearing impaired adults after hearing aid fitting. *Quality of Life Research* 2002;11:299-307.
130. Arrindell W. Changes in waiting-list patients over time: data on some commonly-used measures. Beware! *Behaviour Research and Therapy* 2001;39:1227-1247.

131. Bernhard J, Lowy A, Maibach R, Hürny C, for the Swiss Group for Clinical Cancer Research (SAKK) and the Swiss Institute for Applied Cancer Research (SIAC) B, Switzerland. Response shift in the perception of health for utility evaluation: an explorative investigation. *European Journal of Cancer* 2001;37:1729-1735.
132. Rohs FR, Langone CA, Coleman RK. Response shift bias: a problem in evaluating nutrition training using self-report measures. *Journal of Nutrition Education* 2001;33(3):165-170.
133. Bar-On D, Lazar A, Amir M. Quantitative assessment of response shift in QOL research. *Social Indicators Research* 2000;49:37-49.
134. Jansen S, Stiggelbout A, Nooij M, Noordijk E, Kievit J. Response shift in quality of life measurement in early-stage breast cancer patients undergoing radiotherapy. *Quality of Life Research* 2001;9:603-615.
135. O'Boyle CA, McGee HM, Browne JP. Measuring response shift using the Schedule for Evaluation of Individual Quality of Life. In: Schwartz CE, Sprangers MA, editors. *Adaptation to changing health: response shift in quality-of-life research*. Washington, DC: American Psychological Association, 2000.
136. Postular D, Adang E. Response shift and adaptation in chronically ill patients. *Medical Decision Making* 2000;20:186-193.
137. Pratt CC, McGuigan WM, Katzev AR. Measuring program outcomes: using retrospective pretest methodology. *American Journal of Evaluation* 2000;21(3):341-349.
138. Rapkin BD. Personal goals and response shifts: understanding the impact of illness and events on the quality of life of people living with AIDS. In: Schwartz CE, Sprangers MA, editors. *Adaptation to changing health: response shift in quality-of-life research*. Washington, DC: American Psychological Association, 2000.
139. Sprangers MA, van Dam FS, Broersen J, Lodder L, Wever L, Visser MR, et al. Response shift and fatigue: the use of the thentest approach. In: Schwartz CE, Sprangers MA, editors. *Adaptation to changing health: response shift in quality-of-life research*. Washington, DC: American Psychological Association, 2000.
140. Thompson-Fawcett M, Richard C, O'Connor B, Cohen Z, McLeod R. Quality of life is excellent after a pelvic pouch for colitis-associated neoplasia. *Diseases of the Colon and Rectum* 2000;43(11):1497-1502.
141. Visser MR, Smets EM, Sprangers MA, Haes HJ. How response shift may affect the measurement of change in fatigue. *Journal of Pain and Symptom Management* 2000;20(1):12-18.
142. Daltroy LH, Larson MG, Eaton HM, Phillips CB, Liang MH. Discrepancies between self-reported and observed physical function in the elderly: the influence of response shift and other factors. *Social Science and Medicine* 1999;48:1549-1561.
143. Schwartz CE, Feinberg RG, Jilinskaia E, Applegate JC. An evaluation of a psychosocial intervention for survivors of childhood cancer: paradoxical effects of response shift over time. *Psycho-Oncology* 1999;8:344-354.
144. Sprangers MA, van Dam FS, Broersen J, Lodder L, Wever L, Visser MR, et al. Revealing response shift in longitudinal research on fatigue: the use of the thentest approach. *Acta Oncologica* 1999;38(6):709-718.

145. Manthei RJ. The response-shift bias in a counsellor education programme. *British Journal of Guidance & Counselling* 1997;25(2):229-237.
146. Hickok JT, Morrow GR, McDonald S, Bellg AJ. Frequency and correlates of fatigue in lung cancer patients receiving radiation therapy: implications for management. *Journal of Pain and Symptom Management* 1996;11(6):370-377.
147. Sprangers MAG. Response-shift bias: a challenge to the assessment of patients' quality of life in cancer clinical trials. *Cancer Treatment Reviews* 1996;22(Supplement A):55-62.
148. Goedhart H, Hoogstraten J. The retrospective pretest and the role of pretest information in evaluative studies. *Psychological Reports* 1992;70:699-704.
149. Gutek BA, Winter SJ. Consistency of job satisfaction across situations: fact or framing artifact? *Journal of Vocational Behavior* 1992;41:61-78.
150. Levinson W, Gordon G, Skeff K. Retrospective versus actual pre-course self-assessments. *Evaluation & the health professions* 1990;13(4):445-452.
151. Sprangers M, Hoogstraten J. Pretesting effects in retrospective pretest-posttest designs. *Journal of Applied Psychology* 1989;74(2):265-272.
152. Hoogstraten J, Koele P. A method for analyzing retrospective pretest/posttest designs: II. Application. *Bulletin of the Psychonomic Society* 1988;26(2):124-125.
153. Sprangers M, Hoogstraten J. Response-style effects, response-shift bias, and a bogus-pipeline: a replication. *Psychological Reports* 1988;62:11-16.
154. Zwiebel A. Changing educational counsellors' attitudes toward mental retardation: comparison of two measurement techniques. *International Journal of Rehabilitation Research* 1987;10(4):383-389.
155. Collins LM, Graham JW, Hansen WB, Johnson CA. Agreement between retrospective accounts of substance use and earlier reported substance use. *Applied Psychological Measurement* 1985;9(3):301-309.
156. Hoogstraten J. Influence of objective measures on self-reports in a retrospective pretest-posttest design. *Journal of Experimental Education* 1985;53(4):207-210.
157. Lowy A, Bernhard J. Quantitative assessment of changes in patients' constructs of quality of life: an application of multilevel models. *Quality of Life Research* 2004;13:1177-1185.
158. Ubel PA, Loewenstein G, Jepson C. Whose quality of life? A commentary exploring discrepancies between health state evaluations of patients and the general public. *Quality of Life Research* 2003;12:599-607.
159. Eiser C. Use of quality of life measures in clinical trials. *Ambulatory Pediatrics* 2004;4(4):395-399.
160. Schwartz CE, Sprangers MAG. An introduction to quality of life assessment in oncology: the value of measuring patient-reported outcomes. *The American Journal of Managed Care* 2002;8(Supplement):550-559.

161. Wilson I. Clinical understanding and clinical implications of response shift. *Social Science and Medicine* 1999;48(11):1577-1588.
162. Norman P, Parker S. The interpretation of change in verbal reports: implications for health psychology. *Psychology and Health* 1996;11:301-314.
163. de Haes JCJM, van Knippenberg FCE. The quality of life of cancer patients: a review of the literature. *Social Science and Medicine* 1985;20(8):809-817.
164. Bernhard J, Hürny C, Maibach R, Herrmann R, Laffer U, for the Swiss Group for Clinical Cancer Research (SAKK) and the Swiss Institute for Applied Cancer Research (SAKK) B, Switzerland. Quality of life as subjective experience: reframing of perception in patients with colon cancer undergoing radical resection with or without adjuvant chemotherapy. *Annals of Oncology* 1999;10:775-782.
165. Bray JH, Maxwell SE, Howard GS. Methods of analysis with response-shift bias. *Educational and Psychological Measurement* 1984;44:781-804.
166. Sprangers M, Schwartz C. The challenge of response shift for quality-of-life-based clinical oncology research. *Annals of Oncology* 1999;10:747-749.
167. Cronbach LJ, Furby L. How should we measure "change" - or should we? *Psychological Bulletin* 1970;74(1):68-80.
168. Terborg JR, Howard GS, Maxwell SE. Evaluating planned organisational change: a method for assessing alpha, beta, and gamma change. *Academy of Management Review* 1980;5(1):109-121.
169. Schwartz CE, Sprangers MA. Discussion: implications of response shift for clinical research. In: Schwartz CE, Sprangers MA, editors. *Adaptation to changing health: response shift in quality-of-life research*. Washington, DC: American Psychological Association, 2000.
170. Smets E, Garssen B, Schuster-Uitterhoeve A, de Haes J. Fatigue in cancer patients. *British Journal of Cancer* 1993;68:220-224.
171. Adang E. A comment on "Response shift and adaptation in chronically ill patients" by Postulart and Adang. *Medical Decision Making* 2001;21(5):423-424.
172. Office of Population Censuses and Surveys. *Morbidity statistics from general practice: fourth national study 1991 -1992*. London: HMSO, 1995.
173. Hollingworth W, MacKenzie R, Todd C, Dixon AK. Measuring changes in quality of life following magnetic resonance imaging of the knee: SF-36, EuroQol or Rosser Index? *Quality of Life Research* 1995;4:325-334.
174. Garratt A, Brealey S, Gillespie W, Team icwtDT. Patient-assessed health instruments for the knee: a structured review. *Rheumatology* 2004;43(11):1414-1423.
175. Borsa PA, Lephart SM, Irrgang JJ. Comparison of performance-based and patient-reported measures of function in anterior-cruciate ligament-deficient individuals. *Journal of Orthopaedic and Sports Physical Therapy* 1998;28(6):392-399.
176. Lysholm J, Gillquist J. Evaluation of knee ligament surgery results with special emphasis on use of a scoring scale. *The American Journal of Sports Medicine* 1982;10(3):150-154.

177. Mohtadi N. Development and validation of the Quality of Life Outcome Measure (Questionnaire) for chronic anterior cruciate ligament deficiency. *The American Journal of Sports Medicine* 1998;26(3):350-359.
178. Rockborn P, Hamberg P, Gillquist J. Arthroscopic meniscectomy: treatment costs and postoperative function in a historical perspective. *Acta Orthopaedica Scandinavica* 2000;71(5):455-460.
179. Wexler G, Bach BJ, Bush-Joseph C, Smink D, Ferrari J, Bojchuk J. Outcomes of anterior cruciate ligament reconstruction in patients with Workers' Compensation claims. *Arthroscopy* 2000;16(1):49-58.
180. Watura R, Lloyd DC, Chawda S. Magnetic resonance imaging of the knee: direct access for general practitioners. *British Medical Journal* 1995;311:1614.
181. Robling M, Kinnersley P, Houston H, Hourihan M, Cohen D, Hale J. An exploration of GPs' use of MRI: a critical incident study. *Family Practice* 1998;15(3):236-243.
182. Robling M, Kinnersley P, Houston H, Hourihan M, Cohen D, Hale J. The impact of guidelines and different dissemination strategies on GPs' knowledge of magnetic resonance imaging. *Medical Education* 1999;33:777-779.
183. Robling MR, Houston H, Kinnersley P, Hourihan M, Cohen D, Hale J, et al. General practitioners' use of magnetic resonance imaging: an open randomised trial comparing telephone and written requests and an open randomised controlled trial of different methods of guidelines dissemination. *Clinical Radiology* 2002;57:402-407.
184. Garratt A, Brealey S, Robling M, Atwell C, Russell I. Development and evaluation of the Internal Derangement of the Knee Questionnaire (IDK-38). *Rheumatology*;Submitted for publication.
185. Patton MQ. *Qualitative Research and Evaluation Methods*. Thousand Oaks: Sage Publications Inc., 2002.
186. Ruta DA, Garratt AM, Leng M, Russell IT, MacDonald LM. A new approach to the measurement of quality of life: The Patient Generated Index. *Medical Care* 1994;32(11):1109-1126.
187. Fayers PM, Machin D. *Quality of Life: Assessment, Analysis and Interpretation*: John Wiley & Sons Ltd, 2000.
188. McGee H, O'Boyle CA, Hickey A, O'Malley K, Joyce CRB. Assessing the quality of life of the individual: the SEIQoL with a healthy and gastroenterology unit population. *Psychological Medicine* 1991;21:749-759.
189. Browne JP, O'Boyle CA, McGee HM, McDonald NJ, Joyce C. Development of a direct weighting procedure for quality of life domains. *Quality of Life Research* 1997;6:301-309.
190. Browne J, O'Boyle C, McGee H, Joyce C, McDonald N, O'Malley K, et al. Individual quality of life in the healthy elderly. *Quality of Life Research* 1994;3:235-244.
191. Hickey AM, Bury G, O'Boyle CA, Bradley F, O'Kelly FD, Shannon W. A new short form individual quality of life measure (SEIQoL-DW): application in a cohort of individuals with HIV/AIDS. *British Medical Journal* 1996;313:29-33.

192. O'Boyle C, Browne J, Hickey A, McGee H, CRB J. *Schedule for the Evaluation of Individual Quality of Life (SEIQoL): a direct weighting for quality of life domains (SEIQoL-DW) - administration manual*. Dublin: Department of Psychology, Royal College of Surgeons in Ireland, 1995.
193. Waldron D, O'Boyle CA, Kearney M, Moriarty M, Carney D. Quality-of-life measurement in advanced cancer: assessing the individual. *Journal of Clinical Oncology* 1999;17(11):3603-3611.
194. Patel KK, Veenstra DL, Patrick DL. A review of selected patient-generated outcome measures and their application in clinical trials. *Value in Health* 2003;6(5):595-603.
195. Neudert C, Wasner M, Borasio GD. Patients' assessment of quality of life instruments: a randomised study of SIP, SF-36 and SEIQoL-DW in patients with amyotrophic lateral sclerosis. *Journal of the Neurological Sciences* 2001;191:103-109.
196. Joyce C, Hickey A, McGee H, O'Boyle C. A theory-based method for the evaluation of individual quality of life: The SEIQoL. *Quality of Life Research* 2003;12:275-280.
197. Lindblad ÅK, Ring L, Glimelius B, Hansson MG. Quality of life assessment in oncology. *Acta Oncologica* 2002;41(6):507-516.
198. Macduff C. Respondent-generated quality of life measures: useful tools for nursing or more fool's gold? *Journal of Advanced Nursing* 2000;32(2):375-382.
199. Ruta DA, Garratt AM, Russell IT. Patient centred assessment of quality of life for patients with four common conditions. *Quality in Health Care* 1999;8:22-29.
200. Rees J, O'Boyle C, MacDonagh R. Quality of life: impact of chronic illness on the partner. *Journal of the Royal Society of Medicine* 2001;94:563-566.
201. O'Boyle CA, McGee H, Hickey A, O'Malley K, Joyce C. Individual quality of life in patients undergoing hip replacement. *The Lancet* 1992;339:1088-91.
202. Lintern TC, Beaumont JG, Kenealy PM, Murrell RC. Quality of Life (QoL) in severely disabled multiple sclerosis patients: Comparison of three QoL measures using multidimensional scaling. *Quality of Life Research* 2001;10:371-378.
203. Wettergren L, Björkholm M, Axdorph U, Bowling A, Langius-Eklöf A. Individual quality of life in long-term survivors of Hodgkin's lymphoma - a comparative study. *Quality of Life Research* 2003;12:545-554.
204. Tovbin D, Gidron Y, Jean T, Granovsky R, Schnieder A. Relative importance and intercorrelations between psychosocial factors and individualized quality of life of hemodialysis patients. *Quality of Life Research* 2003;12:709-717.
205. Scheier MF, Carver CS, Bridges MW. Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the Life Orientation Test. *Journal of Personality and Social Psychology* 1994;67(6):1063-1078.
206. Maxwell JA. *Qualitative Research Design: An Interactive Approach*: Sage Publications Inc., 1996.
207. QSR NUD*IST. *User Guide*. London: Sage, 2003.

208. Miles MB, Huberman AM. *Qualitative data analysis: an expanded sourcebook*. Second Edition ed. Thousand Oaks, California: SAGE Publications, Inc., 1994.
209. Silverman D. *Interpreting qualitative data: methods for analysing talk, text and interaction*. 2nd Edition ed. London: Sage Publications, 2001.
210. Seale C. *The quality of qualitative research*. London: Sage Publications, 1999.
211. World Health Organization. Constitution of the World Health Organization. Geneva: World Health Organization, 1948.
212. Garratt A, Schmidt L, Mackintosh A, Fitzpatrick R. Quality of life measurement: bibliographic study of patient assessed health outcome measures. *British Medical Journal* 2002;324:1417.
213. Hunt S. The problem of quality of life. *Quality of Life Research* 1997;6:205-212.
214. Fitzpatrick R, Fletcher A, Gore S, Jones D, Spiegelhalter D, Cox D. Quality of life measures in health care. I: Applications and issues in assessment. *British Medical Journal* 1992;305:1074-1077.
215. Pouchot J. Toward individualized quality-of-life assessment. *Joint Bone Spine* 2000;67:83-5.
216. Day H, Jankey SG. Lessons from the literature: towards a holistic model of quality of life. In: Renwick R, Brown I, Nagler M, editors. *Quality of life in health promotion and rehabilitation*. Thousand Oaks: Sage Publications, 1996.
217. Campbell A. *The sense of well-being in America: Recent patterns and trends*. New York: McGraw-Hill, 1981.
218. Johnson LB. Great Society Speech, University of Michigan, 1964.
219. Calman K. Quality of life in cancer patients - an hypothesis. *Journal of Medical Ethics* 1984;10:124-127.
220. Farquhar M. Definitions of quality of life: a taxonomy. *Journal of Advanced Nursing* 1995;22:502-508.
221. Fitzpatrick R. Assessment of quality of life as an outcome: finding measurements that reflect individual's priorities [Editorial]. *Quality in Health Care* 1999;8:1-2.
222. Carr AJ, Higginson IJ. Are quality of life measures patient centred? *British Medical Journal* 2001;322:1357-1360.
223. Bullinger M. Cognitive theories and individual quality of life assessment. In: Joyce C, McGee HM, O'Boyle CA, editors. *Individual quality of life: approaches to conceptualisation and assessment*. First edition ed. Amsterdam: Harwood Academic Publishers, 1999:29-39.
224. Kaasa S, Loge JH. Quality of life in palliative medicine - principles and practice. In: Doyle D, Hanks G, Cherny N, Calman K, editors. *Oxford Textbook of Palliative Medicine*. Oxford: Oxford University Press, 2004.
225. Ryff C, Singer BH, Love GD. Positive health: connecting well-being with biology. *Philosophical transactions of The Royal Society of London. Series B: Biological Sciences* 2004;359:1383-1394.
226. Broadie S, Rowe C. *Aristotle: Nicomachean Ethics. Translation, introduction and commentary*. Oxford: Oxford University Press, 2002.
227. Meeberg GA. Quality of life: a concept analysis. *Journal of Advanced Nursing* 1993;18:32-38.

228. Haas BK. A multidisciplinary concept analysis of quality of life. *Western Journal of Nursing Research* 1999;21(6):728-742.
229. Browne JP, McGee HM, O'Boyle CA. Conceptual approaches to the assessment of quality of life. *Psychology and Health* 1997;12:737-751.
230. Allison PJ, Locker D, Feine JS. Quality of Life: A Dynamic Construct. *Social Science and Medicine* 1997;45(2):221-230.
231. Rosenberg R. Health-related quality of life between naturalism and hermeneutics. *Social Science and Medicine* 1995;41(10):1411-1415.
232. Gill TM, Feinstein AR. A critical appraisal of the quality of quality-of-life measurements. *Journal of The American Medical Association* 1994;272(8):619-626.
233. Hickey A, O'Boyle CA, McGee H, Joyce CRB. The Schedule for the Evaluation of Individual Quality of Life. In: Joyce CRB, O'Boyle CA, McGee H, editors. *Individual quality of life: approaches to conceptualisation and assessment*. First ed. Amsterdam: Harwood Academic Publishers, 1999.
234. Sevenhuysen GP, Trumble-Waddell J. A new perspective on quality of life. *Journal of Clinical Epidemiology* 1997;50(3):231-232.
235. O'Boyle CA, Waldron D. Quality of life issues in palliative medicine. *Journal of Neurology* 1997;244(Supplement 4):S18-S25.
236. Campbell S, Whyte F. The quality of life of cancer patients participating in phase I clinical trials using SEIQoL-DW. *Journal of Advanced Nursing* 1999;30(2):335-343.
237. Sprangers MA. Quality-of-life assessment in oncology: achievements and challenges. *Acta Oncologica* 2002;41(3):229-237.
238. Fallowfield L. Participation of patients in decisions about treatment for cancer. *British Medical Journal* 2001;323:1144.
239. Johnson DS, Smith RB. Outcome measurement in the ACL deficient knee - what's the score? *The Knee* 2000;8:51-57.
240. Editorial. Quality of life. *The Lancet* 1991;338:350-351.
241. Hadorn DC, The Steering Committee of the Western Canada Waiting List Project. Setting priorities for waiting lists: defining our terms. *Canadian Medical Association Journal* 2000;163(7):857-860.
242. Lubeck DP. Health-related quality of life measurements and studies in rheumatoid arthritis. *The American Journal of Managed Care* 2002;8(9):811-820.
243. Neudert C, Wasner M, Borasio GD. Individual quality of life is not correlated with health-related quality of life or physical function in patients with amyotrophic lateral sclerosis. *Journal of Palliative Medicine* 2004;7(4):551-557.
244. Greenhalgh J, Meadows K. The effectiveness of the use of patient-based measures of health in routine practice in improving the process and outcomes of patient care: a literature review. *Journal of Evaluation in Clinical Practice* 1999;5(4):401-416.
245. Sepúlveda C, Marlin A, Yoshida T, Ullrich A. Palliative Care: The World Health Organization's global perspective. *Journal of Pain and Symptom Management* 2002;24(2):91-96.
246. Doyle D, Hanks G, Cherny NI, Calman K. Introduction. In: Doyle D, Hanks G, Cherny N, Calman K, editors. *Oxford Textbook of Palliative Medicine*. Oxford: Oxford University Press, 2004.

247. Lubeck D. Patient-reported outcomes and their role in the assessment of rheumatoid arthritis. *Pharmacoeconomics* 2004;22(Supplement 1):27-38.
248. Russack SM, Croft jr JD, Furst DE, Hohlbauch A, Liang MH, Moreland L, et al. The use of rheumatoid arthritis health-related quality of life patient questionnaires in clinical practice: lessons learned. *Arthritis & Rheumatism (Arthritis Care & Research)* 2003;49(4):574-584.
249. Strauss A, Corbin J. *Basics of Qualitative Research*. 2nd Edition ed. Thousand Oaks: Sage Publications, 1998.
250. Coffey A, Holbrook B, Atkinson P. Qualitative data analysis: technologies and representations. *Sociological Research Online* 1996;1(1):4.
251. Tegner Y, Lysholm J. Rating systems in the evaluation of knee ligament injuries. *Clinical Orthopaedics and Related Research* 1985;198:43-49.
252. Irrgang JJ, Anderson AF. Development and validation of health-related quality of life measures for the knee. *Clinical Orthopaedics and Related Research* 2002;402:95-109.
253. Marx RG. Knee rating scales. *Arthroscopy: The Journal of Arthroscopic and Related Surgery* 2003;19(10):1103-1108.
254. Leigh Brown AP, Kennedy ADM, Grant AM, Campbell J, Macnicol MF, Torgerson DJ. The development and validation of the Edinburgh Knee Function Scale: a simple tool for outcome measurement in non-surgical patients. *The Knee* 1999;6:115-123.
255. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee injury and Osteoarthritis Outcome Score (KOOS) - development of a self-administered outcome measure. *Journal of Orthopaedic and Sports Physical Therapy* 1998;78(2):88-96.
256. Bayle B, Kemoun G, Migaud H, Thevenon A. Comparison of two modes of administration of a personalised quality of life scale in a longitudinal study of hip arthroplasty. *Joint Bone Spine* 2000;67:101-106.
257. Fayers PM, Hand DJ. Causal variables, indicator variables and measurement scales: an example from quality of life. *Journal of the Royal Statistical Society, Series A* 2002;165:233-261.
258. Rawles J. Catigating QALYs. *Journal of Medical Ethics* 1989;15:143-147.
259. McCall S. Quality of Life. *Social Indicators Research* 1975;2:229-248.
260. Collins D. Pretesting survey instruments: an overview of cognitive methods. *Quality of Life Research* 2003;12:229-238.
261. Wagner J, Abbott G, Lett S. Age related differences in individual quality of life domains in youth with type 1 diabetes. *Health and Quality of Life Outcomes* 2004;2(1):54.
262. Sprangers MA, Schwartz CE. Discussion: methodological pathways. In: Schwartz CE, Sprangers MA, editors. *Adaptation to changing health: response shift in quality-of-life research*. Washington, DC: American Psychological Association, 2000.
263. Smout S, Koudstaal PJ, Ribbers GM, Janssen WGM, Passchier J. Struck by stroke: a pilot study exploring quality of life and coping patterns in younger patients and spouses. *International Journal of Rehabilitation Research* 2001;24:261-268.
264. Michalos A. Multiple discrepancies theory (MDT). *Social Indicators Research* 1985;16:347-413.

265. Llewellyn-Thomas HA, Schwartz CE. Response shift effects on patients' evaluations of health states: sources of artifact. In: Schwartz CE, Sprangers MA, editors. *Adaptation to changing health: response shift in quality-of-life research*. Washington, DC: American Psychological Association, 2000.
266. Bowling A, Gabriel Z, Dykes J, Marriot Dowding L, Evans O, Fleissig A, et al. Let's ask them: a national survey of definitions of quality of life and its enhancement among people aged 65 and over. *International Journal of Aging and Human Development* 2003;56(4):269-306.
267. Herrmann D. Reporting current, past, and changed health status. What we know about distortion. *Medical Care* 1995;33(4 Supplement):AS89-94.
268. Visser MRM, Oort FJ, Sprangers MA. Methods to detect response shift in quality of life data: a convergent validity study. *Quality of Life Research* 2005;14:629-639.
269. Fransen M, Edmonds J. Reliability and validity of the EuroQol in patients with osteoarthritis of the knee. *Rheumatology* 1999;38:807-813.
270. Hurst NP, Kind P, Ruta D, Hunter M, A S. Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D). *British Journal of Rheumatology* 1997;36:551-559.
271. Coons SJ, Rao S, Keininger DL, Hays RD. A comparative review of generic quality-of-life instruments. *Pharmacoeconomics* 2000;17(1):13-35.
272. Siegel S, Castellan Jr NJ. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill Book Company, 1988.
273. Goodwin PC, Morrissey MC, Omar RZ, Brown M, Southall K, McAuliffe TB. Effectiveness of supervised physical therapy in the early period after arthroscopic partial meniscectomy. *Physical Therapy* 2003;83(6):520-535.
274. Aristotle. *The Nicomachean Ethics*. London: Penguin, 2004.
275. Bromberg MB, Forshew DA. Comparison of instruments addressing quality of life in patients with ALS and their caregivers. *Neurology* 2002;58:320-322.
276. Mountain LA, Campbell SE, Seymour DG, Primrose WR, Whyte MI. Assessment of individual quality of life using the SEIQoL-DW in older medical patients. *Q J Med* 2004;97:519-524.
277. Frick E, Borasio G, Zehentner H, Fischer N, Bumeder I. Individual quality of life of patients undergoing autologous peripheral blood stem cell transplantation. *Psycho-Oncology* 2004;13:116-124.
278. Clarke S, Hickey A, O'Boyle C, Hardiman O. Assessing individual quality of life in amyotrophic lateral sclerosis. *Quality of Life Research* 2001;10:149-158.
279. Prince PN, Gerber GJ. Measuring subjective quality of life in people with serious mental illness using the SEIQoL-DW. *Quality of Life Research* 2001;10:117-122.
280. Altman DG. *Practical statistics for medical research*. London: Chapman and Hall, 1991.
281. Norman GR, Streiner DL. *Biostatistics: The Bare Essentials*. Hamilton: B. C. Decker Inc., 2000.

282. Ramström H, Erwander I, Mared L, Kornfält R, Seiving B. Pharmaceutical intervention in the care of cystic fibrosis patients. *Journal of Clinical Pharmacy and Therapeutics* 2000;25:427-434.
283. Bowling A. What things are important in people's lives? A survey of the public's judgements to inform scales of health related quality of life. *Social Science and Medicine* 1995;41(10):1447-1462.
284. Gribbin GM, Kenny RA, McCue P, Toff WD, Bexton RS, McComb JM. Individualised quality of life after pacing. Does mode matter? *Europace* 2004;6:552-560.
285. Neudert C, Clarke S, Hickey A, O'Boyle C, Borasio GD. Existential factors are an important determinant of quality of life in amyotrophic lateral sclerosis. *Journal of Neurology* 2002;249(Supplement 1):S23.
286. Montgomery C, Pocock M, Titley K, Lloyd K. Individual quality of life in patients with leukaemia and lymphoma. *Psycho-Oncology* 2002;11:239-243.
287. Howell J, Handoll H. Surgical treatment for meniscal injuries of the knee in adults (Review). *The Cochrane Database of Systematic Reviews*, 1999:Issue 1. Art. No.: CD001353. DOI: 10.1002/14651858.CD001353.
288. Peccin M, Almeida G, Amaro J, Cohen M, Soares B, Atallah A. Interventions for treating posterior cruciate ligament injuries of the knee in adults (Review). *The Cochrane Database of Systematic Reviews*, 2005:Issue 2. Art. No.: CD002939.pub2. DOI: 10.1002/14651858.CD002939.pub2.
289. Linko E, Harilainen A, Malmivaara A, Seitsalo S. Surgical versus conservative interventions for anterior cruciate ligament ruptures in adults (Review). *The Cochrane Database of Systematic Reviews*, 2005:Issue 2. Art. No.: CD001356.pub3. DOI: 10.1002/14651858.CD001356.pub3.
290. Marx RG, Jones EC, Angel M, Wickiewicz TL, Warren RF. Beliefs and attitudes of members of the American Academy of Orthopaedic Surgeons regarding the treatment of anterior cruciate ligament injury. *Arthroscopy: The Journal of Arthroscopic and Related Surgery* 2003;19(7):762-770.
291. Dunn W, Lyman S, Lincoln AE, Amoroso PJ, Wickiewicz T, Marx RG. The effect of anterior cruciate ligament reconstruction on the risk of knee injury. *The American Journal of Sports Medicine* 2004;32(8):1906-1914.
292. Rath E, Richmond JC. The menisci: basic science and advances in treatment. *British Journal of Sports Medicine* 2000;34:252-257.
293. Venkatachalam S, Godsiff S, Harding M. Review of the clinical results of arthroscopic meniscal repair. *The Knee* 2001;8:129-133.
294. de Roeck NJ, Lang-Stevenson A. Meniscal tears sustained awaiting anterior cruciate ligament reconstruction. *Injury* 2003;34:343-345.
295. O'Conner DP, Laughlin MS, Woods GW. Factors related to additional knee injuries after anterior cruciate ligament injury. *Arthroscopy: The Journal of Arthroscopic and Related Surgery* 2005;21(4):431-438.
296. NHS Wales. <http://www.hsw.wales.nhs.uk/ipd/homepage.htm>
297. Williams RL, Williams LA, Watura R, Fairclough JA. Impact of MRI on a knee arthroscopy waiting list. *Annals of the Royal College of Surgeons* 1996;78:450-452.

298. Perkins T. Waiting times for MRI. Personal communication., 2005.
299. Lewis S, Barer ML, Sanmartin C, Sheps S, Shortt SED, McDonald PW. Ending waiting-list mismanagement: principles and practice. *Canadian Medical Association Journal* 2000;162(9):1297-1300.
300. Kelly KD. Equity in waiting times for major joint arthroplasty. *Canadian Journal of Surgery* 2002;45(4):269-276.
301. Llewellyn-Thomas HA, Arshinoff R, Bell M, Williams JI, Naylor CD. In the queue for total joint replacement: patients' perspectives on waiting times. *Journal of Evaluation in Clinical Practice* 1998;4(1):63-74.
302. Dunn E, Black C, Alonso J, Norregaard JC, Anderson GF. Patients' acceptance of waiting for cataract surgery: what makes a wait too long? *Social Science and Medicine* 1997;44(11):1603-1610.
303. Sanders C, Donovan JL, Dieppe PA. Unmet need for joint replacement: a qualitative investigation of barriers to treatment among individuals with severe pain and disability of the hip and knee. *Rheumatology* 2004;43:353-357.
304. Derrett S, Paul C, Morris J. Waiting for elective surgery: effects on health-related quality of life. *International Journal for Quality in Health Care* 1999;11(1):47-57.
305. Kelly KD, Voaklander D, Kramer G, Johnston DWC, Redfern L, Suarez-Almazor ME. The impact of health status on waiting time for major joint arthroplasty. *The Journal of Arthroplasty* 2000;15(7):877-883.
306. Sanmartin C, Shortt SED, Barer ML, Sheps S, Lewis S, McDonald PW. Waiting for medical services in Canada: lots of heat, but little light. *Canadian Medical Association Journal* 2000;162(9):1305-1310.
307. Hadorn DC, Holmes AC. The New Zealand priority criteria project. Part 1: Overview. *British Medical Journal* 1997;314(7074):131-134.
308. Derrett S, Paul C, Herbison P, Williams H. Evaluation of explicit prioritisation for elective surgery: a prospective study. *Journal of Health Services Research and Policy* 2002;7(Supplement 1):S1:14-22.
309. Edwards RT, Boland A, Wilkinson C, Cohen D, Williams J. Clinical and lay preferences for the explicit prioritisation of elective waiting lists: survey evidence from Wales. *Health Policy* 2003;63:229-237.
310. Ostendorf M, Buskens E, van Stel H, Schrijvers A, Marting L, Dhert W, et al. Waiting for total hip arthroplasty: avoidable loss in quality time and preventable deterioration. *The Journal of Arthroplasty* 2004;19(3):302-309.
311. Peters TJ, Sanders C, Dieppe P, Donovan J. Factors associated with change in pain and disability over time: a community-based prospective observational study of hip and knee osteoarthritis. *British Journal of General Practice* 2005;55(512):205-211.
312. Mahon JL, Bourne RB, Rorabeck CH, Feeny DH, Stitt L, Webster-Bogaert S. Health-related quality of life and mobility of patients awaiting elective total hip arthroplasty: a prospective study. *Canadian Medical Association Journal* 2002;167(10):1115-1121.
313. Kelly KD, Voaklander DC, Johnston DWC, Newman SC, Suarez-Almazor ME. Change in pain and function while waiting for major joint arthroplasty. *The Journal of Arthroplasty* 2001;16(3):351-359.

314. Stewart M, Brown JB, Weston WW, McWhinney IR, McWilliam CL, Freeman TR. *Patient-centred medicine: Transforming the clinical method*. First ed. Oxford: Radcliffe Medical Press, 1995.
315. Beaton DE. Understanding the relevance of measured change through studies of responsiveness. *Spine* 2000;25(24):3192-3199.
316. Fischer D, Stewart AL, Bloch DA, Lorig K, Laurent D, Holman H. Capturing the patient's view of change as a clinical outcome measure. *Journal of The American Medical Association* 1999;282(12):1157-1162.
317. Carr A. Problems in measuring or interpreting change in patient outcomes. *Osteoarthritis and Cartilage* 2002;10:503-505.
318. Mick DJ, Ackerman MH. New perspectives on advanced practice nursing case management for aging patients. *Critical Care Nursing Clinics of North America* 2002;14:281-291.
319. Parducci A. An historical introduction to response shift. In: Schwartz CE, Sprangers MA, editors. *Adaptation to changing health: response shift in quality-of-life research*. Washington, DC: American Psychological Association, 2000.
320. Breetvelt I, van Dam F. Underreporting by cancer patients: the case of response-shift. *Social Science and Medicine* 1991;32(9):981-987.
321. Koele P, Hoogstraten J. A method for analyzing retrospective pretest/posttest designs: I. Theory. *Bulletin of the Psychonomic Society* 1988;26(1):51-54.
322. Sprangers M. Subject bias and the retrospective pretest in retrospect. *Bulletin of the Psychonomic Society* 1989;27(1):11-14.