Can clinical presentation predict response to a non-surgical chronic disease management program for hip and knee osteoarthritis?

Jillian Eyles BAppSc(Phty)

A Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

Northern Clinical School, Sydney Medical School

Faculty of Medicine and Health

The University of Sydney

2019

Supervisors' Statement

As supervisors of Jillian Eyles' doctoral work, we certify that we consider her Thesis "Can clinical presentation predict response to a non-surgical chronic disease management program for hip and knee osteoarthritis?" sufficiently well presented to be examined and certify that it does not exceed the prescribed word limit or any extended word limit for which prior approval has been granted.

Professor David J Hunter

Institute of Bone and Joint Research, The Kolling Institute of Medical Research,

Northern Clinical School, Sydney Medical School, Faculty of Medicine and Health,

The University of Sydney, Australia, Rheumatology Department, Royal North Shore Hospital

Date 1st March, 2019

Dr Kathryn Mills

Faculty of Medicine and Health Sciences

Macquarie University, Sydney, Australia

Date 1st March, 2019

Dr Barbara R Lucas

Physiotherapy Department, Royal North Shore Hospital,

The Kolling Institute of Medical Research, John Walsh Centre for Rehabilitation Research,

Discipline of Child and Adolescent Health, Sydney Medical School, Faculty of Medicine and

Health, The University of Sydney, Australia

Date 1st March, 2019

Candidate's Statement

I, Jillian Eyles, hereby declare that this submission is my own work and that it contains no material previously published or written by another person except where acknowledged in the text. Nor does it contain material which has been accepted for the award of another degree.

I, Jillian Eyles, understand that if I am awarded a higher degree for my Thesis entitled "Can clinical presentation predict response to a non-surgical chronic disease management program for hip and knee osteoarthritis?" being lodged herewith for examination, the Thesis will be lodged in the University library and be available immediately for use. I agree that the University Librarian (or in the case of a department, the Head of the Department) may supply a photocopy or microform of the Thesis to an individual for research or study or to a library.

Date 1st March, 2019

Table of contents

Tables

195

Figures

Acknowledgements

Christopher, I cannot thank you enough for the love and support you have tirelessly shown me, especially over the past six years of my candidature. I am blessed that you are my best friend, an exceptional partner in life and a wonderful father to our children. Despite both of us leading extremely busy lives, we still manage to have amazing adventures together. I pledge to you that the next adventure will not be another PhD! I love you, I can't wait to see what we are going to do next. To my extraordinary children Brayden and Lachlan. You are both clever, funny and kind, therefore I am the luckiest mother in the world. You were so young when I started on this journey (Lachlan three, Brayden five) you have probably forgotten what it is like to have a mother who is not a 'crazy PhD Mum'. Thank you both for your patience and understanding when I was sitting in front of the computer day, night and every weekend. These last few months you have both offered me endless support and kindness for which I will always be grateful. I love you both so much and look to enjoying our journey together as you continue to grow and learn. Never forget that you can do anything- reach for the stars!

Mum and Dad, you have always been my greatest advocates and have always told me that I can do anything. Well, look what happened! Thank you for your abundant love, support and encouragement. I love and respect you both so much. I also love and respect my beautiful, brave sister Alison and gorgeous, strong niece Isabelle, your support has been so important to me. The last few years have been extremely tough for you both. I miss Dave terribly, but I

can't even begin to imagine how much you both miss him. To Marie, Daryl and the extended Large/Eyles clan, thank you for your continued love and support.

David, Kat and Barb I am so appreciative of your guidance during my PhD candidature. Your support and mentorship have been exceptional. David, you are a great leader, it is such an honour to work with you and I can't thank you enough for the many, many opportunities you have offered me. You make me believe that anything is possible. Kat, you are an amazing mentor. I enjoy getting over-excited about study design, outcome measures and statistics with you immensely, but our friendship is way more fun! Barb, you believed in me from the very beginning and advocated strongly that I should embark on this PhD adventure. Thank you for being a wonderful friend, mentor and colleague.

To my fantastic friends Joce, Manuela Sarah and Matt, I could not have done this without you, you have all been there for me, every day, especially in the last few months. I am so grateful for your friendship and support. Bec, Lisa and Jess, you have all exercised with me to help keep me sane and listened to me when I needed you to. Thank you for your precious friendships and for being my fitness buddies. To my fabulous group of friends- you know who you are- thank you for believing in me and tolerating my limited social contact for the last six years. I am so grateful you are still talking to me! And I look forward to spending more time with you. Lastly, my colleagues, many of you are also dear friends to me. You have seen my every day, celebrated my successes, commiserated with me in my defeats, encouraged and supported me to the last through this time. Thank you all.

Publications and Presentations

Published works from this Thesis:

1. **Eyles, JP**, Hunter, DJ, Meneses, S, Collins, N, Dobson, F, Lucas, BR, Mills, K. Instruments assessing attitudes toward or capability regarding self-management of osteoarthritis: a systematic review of measurement properties. Osteoarthritis and Cartilage 2017;25(8):1210-22.
2. **Eyles JP,** Mills K, Lucas BR, Williams MJ, Makovey J, Teoh L, et al. Can we predict those with osteoarthritis who will worsen following a chronic disease management program? Arthritis Care Res (Hoboken). 2016;68(9):1268-77.
3. **Eyles JP,** Lucas BR, Patterson JA, Williams MJ, Weeks K, Fransen M, et al. Does clinical presentation predict response to a nonsurgical chronic disease management program for endstage hip and knee osteoarthritis? J Rheumatol. 2014;41(11):2223-31.
4. **Eyles J**, Lucas BR, Hunter DJ. Targeting care: tailoring nonsurgical management according to clinical presentation. Rheum Dis Clin North Am. 2013;39(1):213-33.

Work from this thesis submitted for publication

1. **Eyles, JP**, Ferreira, M Mills, K, Lucas, BR Robbins, SR Williams, M, Lee, H Appleton, S, Hunter, DJ. Is the Patient Activation Measure a valid measure of osteoarthritis self-management attitudes and capabilities? Results of a Rasch analysis. BMC Health Quality of Life Outcomes. Submitted 7th July, 2018.

Co-authored works arising from osteoarthritis research during the period of candidature:

1. Mills K, **Eyles JP**, Martin MA, Hancock MJ, Hunter DJ. Exploratory study of 6-month pain trajectories in individuals with predominant patellofemoral osteoarthritis: a cohort study. JOPST. 2018:1-38.
2. Leech RD, **Eyles J**, Batt ME, Hunter DJ. Lower extremity osteoarthritis: optimising musculoskeletal health is a growing global concern: a narrative review. BJSM 2018. (in press)
3. Riordan E, Robbins S, Deveza L, Duong V, Oo WM, Wajon A, et al. Radial subluxation in relation to hand strength and radiographic severity in trapeziometacarpal osteoarthritis. Osteoarthritis Cartilage. 2018.(in press)
4. Duong V, Bennell KL, Deveza LA, **Eyles JP,** Hodges PW, Holden MA, et al. Attitudes, beliefs and common practices of hand therapists for base of thumb osteoarthritis in Australia (The ABC Thumb Study). Hand Therapy. 2017;23(1):19-27.
5. Liu X, Machado GC, **Eyles JP**, Ravi V, Hunter DJ. Dietary supplements for treating osteoarthritis: a systematic review and meta-analysis. Br J Sports Med. 2018;52(3):167-75.
6. Paterson KL, Hunter DJ, Metcalf BR, **Eyles J**, Duong V, Kazsa J, et al. Efficacy of intra-articular injections of platelet-rich plasma as a symptom- and disease-modifying treatment for knee osteoarthritis - the RESTORE trial protocol. BMC musculoskeletal disorders. 2018;19(1):272.
7. Liu X, **Eyles J**, McLachlan AJ, Mobasheri A. Which supplements can I recommend to my osteoarthritis patients? Rheumatology (Oxford, England). 2018;57(suppl_4):iv75-iv87.

8. Castro-Dominguez F, Melo L, **Eyles JP.** Models of healthcare delivery for osteoarthritis. Reumatologia clinica. 2018 (in press).
9. Melo L, Schrieber L, **Eyles J**, Deveza LA, Meneses SRF, Hunter DJ. Comparison of physical examination performance of medical students trained by musculoskeletal versus non-musculoskeletal specialists. Int J Rheum Dis. 2017;20(4):451-9.
10. Deveza LA, Hunter DJ, Wajon A, Bennell KL, Vicenzino B, Hodges P, et al. Efficacy of combined conservative therapies on clinical outcomes in patients with thumb base osteoarthritis: protocol for a randomised, controlled trial (COMBO). BMJ Open. 2017;7(1)
11. Zhang X, **Eyles JP**, Makovey J, Williams MJ, Hunter DJ. Is the effectiveness of patellofemoral bracing modified by patellofemoral alignment and trochlear morphology? BMC Musculoskelet Disord. 2017;18(1):168.
12. Teoh LSG, **Eyles JP**, Makovey J, Williams M, Kwoh CK, Hunter DJ. Observational study of the impact of an individualized multidisciplinary chronic care program for hip and knee osteoarthritis treatment on willingness for surgery. Int J Rheum Dis. 2017;20(10):1383-92.
13. Murphy NJ, **Eyles J**, Bennell KL, Bohensky M, Burns A, Callaghan FM, et al. Protocol for a multi-centre randomised controlled trial comparing arthroscopic hip surgery to physiotherapy-led care for femoroacetabular impingement (FAI): the Australian FASHIoN trial. BMC Musculoskelet Disord. 2017;18(1):406

14. Mills KA, Naylor JM, **Eyles JP**, Roos EM, Hunter DJ. Examining the Minimal Important Difference of Patient-reported Outcome Measures for Individuals with Knee Osteoarthritis: A Model Using the Knee Injury and Osteoarthritis Outcome Score. J Rheumatol. 2016;43(2):395-404.
15. Murphy NJ, **Eyles JP**, Hunter DJ. Hip Osteoarthritis: Etiopathogenesis and Implications for Management. Advances in therapy. 2016;33(11):1921-46.
16. Yu SP, Williams M, **Eyles JP**, Chen JS, Makovey J, Hunter DJ. Effectiveness of knee bracing in osteoarthritis: pragmatic trial in a multidisciplinary clinic. Int J Rheum Dis. 2016;19(3):279-86.

Presentations of work arising from this Thesis

Invited international presentations

1. Eyles JP, Novel models of service delivery for osteoarthritis management. Invited presentation at OARSI Pre-congress workshop: Update on Osteoarthritis Management Programs. Liverpool, UK April 2018.

2. Eyles JP, Measurement of outcomes and their evaluation. Invited presentation at OARSI Pre-congress workshop: Osteoarthritis chronic disease management programs: implementing best practice. Amsterdam, Netherlands April 2016.

International meetings

1. Eyles JP, Mills K, Lucas BR, Robbins SR, O'Connell RL, Williams M, Lee H, Appleton S, Hunter DJ. Is patient activation associated with changes in symptoms following an osteoarthritis management program? Osteoarthritis Research International Congress Accepted for presentation April 2019.
2. Eyles JP, Ferreira M, Mills K, Lucas BR, Robbins SR, Williams M, et al. Does the patient activation measure provide a meaningful measure of OA self-management? Osteoarthritis and Cartilage. 2018;26:S235-S6. Osteoarthritis Research International Congress April 2018, Liverpool UK.
3. Eyles JP, Hunter DJ, Meneses S, Collins N, Dobson F, Lucas B, et al. Measurement Properties of Instruments Assessing Attitudes and Capabilities Regarding Osteoarthritis Self-management: A Systematic Review. Osteoarthritis and Cartilage. 2017;25:S347. Osteoarthritis Research International Congress April 2017, Las Vegas USA.
4. Eyles JP, Mills K, Lucas BR, Williams MJ, Makovey J, Teoh L, et al. Can we predict those who report worsening despite participation in a programme based on OARSI guidelines for non-surgical management of hip and knee OA? Osteoarthritis and Cartilage. 2016;24:S474-S5. Osteoarthritis Research International Congress April 2016, Amsterdam Netherlands.

National meetings

1. Eyles JP, Mills K, Lucas BR, Robbins SR, O'Connell RL, Williams M, Lee H, Appleton S, Hunter DJ. Examining patient activation as a predictor of outcomes following the OACCP. Musculoskeletal Network Forum, Royal Prince Alfred Hospital, Sydney February 2019.
2. **Eyles JP,** Hunter DJ, Meneses S, Collins N, Dobson F, Lucas BR, Mills K. Instruments assessing attitudes towards and/or capabilities regarding self-management of Osteoarthritis; a systematic review of measurement properties. Sydney Musculoskeletal, Bone & Joint Health Alliance Scientific meeting. Sydney October 2017.
3. **Eyles JP**, Lucas BR, McConnell R, Williams MJ, Hunter DJ. Do symptoms of depression moderate the association between weight loss and response of overweight participants to a chronic care programme for hip & knee OA? Momentum 2017 Physiotherapy Conference, ICC Sydney, October 2017.
4. **Eyles JP,** Hunter DJ, Meneses S, Collins N, Dobson F, Lucas BR, Mills K. Which instrument best measures patient attitudes and capabilities regarding osteoarthritis self-management? Momentum 2017 Physiotherapy Conference, ICC Sydney, October 2017.
5. **Eyles JP**, Hunter DJ, Meneses S, Collins N, Dobson F, Lucas BR, Mills K. Instruments assessing attitudes toward or capability regarding self-management of osteoarthritis: a systematic review of measurement properties. New Horizons Conference, UTS Sydney, 2016.

6. **Eyles JP**, Hunter DJ, Meneses S, Collins N, Dobson F, Lucas BR, Mills K. Instruments assessing attitudes towards and/or capabilities regarding self-management of Osteoarthritis- a systematic review of measurement properties. Allied Health Forum, RNSH, Sydney 2016.
7. **Eyles JP**, Mills K, Handcock M, Martin M, Hunter DJ. Can patients with patellofemoral osteoarthritis be sub-grouped at baseline? Journal of Science & Medicine in Sport. December 2015 Volume 19, Supplement, Page e86. Sports Medicine Australia conference Gold Coast Qld, October 2015
8. **Eyles JP**, Lucas BR, Patterson JA, Williams MJ, Weeks K, Fransen M, Hunter DJ. 'Does Clinical Presentation Predict Response to a Nonsurgical Chronic Disease Management Program for Endstage Hip and Knee Osteoarthritis?' 2014 Osteoarthritis Forum at the Kolling Institute of Medical Research, Sydney.
9. Eyles J, Mills K, Handcock M, Martin M, Hunter D. Can patients with patellofemoral osteoarthritis be sub-grouped at baseline? Journal of Science and Medicine in Sport. 2015;19:e86. Sports Medicine Australia Conference, October 2015.

Abstract

Osteoarthritis (OA) is a leading cause of global disability. International guidelines make clear recommendations for evidence-based OA management. However, there is considerable discrepancy between these recommendations and the actual care received by patients. Osteoarthritis management programs (OAMPs) aim to address this evidence-practice gap. There is evidence that some participants improve in pain and function following OAMPs, however, others fail to accomplish these gains. The ability to predict patient outcomes would enable targeting these programs at those people most likely to demonstrate improvement. This Thesis addresses the question: 'Can clinical presentation predict response to an OAMP?' Five studies were conducted to address this question.

Two longitudinal cohort studies were conducted to examine the relationships between participant characteristics and changes in pain and function following 26 weeks of an OAMP. Significant predictors of response were found to include: sex; knee as treatment joint (vs hip); and total joint arthroplasty (TJR) waitlist status. However, the regression models used were not sufficiently sensitive to correctly classify 'responders' or 'worseners'.

We were also interested in examining patients' attitudes and capabilities towards OA self-management as a construct that could potentially predict OAMP outcomes. Before we examined these relationships, we conducted a systematic review (third study) that aimed to identify the instrument assessing OA self-management attitudes and capabilities with the

"best" measurement properties. From this review, little extant measurement property evidence was found to recommend any instrument.

The fourth study examined the measurement properties of the Patient Activation Measure (PAM-13) based on its good face validity. We conducted a cross-sectional cohort study that provided evidence of adequate measurement properties to support the use of PAM-13 to measure OA self-management capabilities in this population. The fifth study examined the relationships between PAM-13 and changes in pain and function. Surprisingly, the PAM-13 scores were not associated with changes in pain or function following 12 or 26 weeks of the OAMP.

It is difficult to determine who will improve or worsen in an OAMP. It does not seem to be based on participants' self-management attitudes or capabilities, although this could be due to the measure we used rather than the construct. Variables found to be significantly associated with outcomes were Timed-Up-and-Go and employment status. Although these findings were statistically significant, the evidence from a single clinical cohort study is insufficient to translate into clinical practice. Hence, these findings should be replicated in larger cohorts.

Abbreviations

ACI Agency for Clinical Innovation

ANOVA analysis of variance

AQoL-6D Assessment of Quality of Life 6 dimensions

AQoL-6D Assessment of Quality of Life 6 dimension version

AS Ankylosing spondylitis

ASES Arthritis Self Efficacy Score

BLOKS Boston-Leeds Osteoarthritis of the Knee Score

BMI Body Mass Index

BML bone marrow lesions

CCM chronic care model

CFI comparative fit index,

CHLC Chance Health Locus of Control

CI confidence interval

COSMIN COnsensus-based Standards for the selection of health Measurement Instruments

COX-2 inhibitors Cyco-Oxygenase-2 inhibitors

CP Controllability for physicians,

CPh Controllability for pharmacists

CSI intra-articular corticosteroid

CTT Classical Test Theory

CTT Classical Test Theory

DASS 21 Depression Anxiety Stress 21 Scale

DASS-21 Depression, Anxiety and Stress Scale 21 item version

df degrees of freedom

DIF Differential item functioning

DIF Differential Item Functioning

EC-17 Effective Consumer Scale

ENAT Educational needs assessment

ESCAPE-knee pain Enabling Self-Management and Coping of Arthritic Knee Pain Through Exercise

FA Factor Analysis,

FA Factor Analysis

FM Fibromyalgia

GFI Goodness of fit index,

heiQ Health Education Impact Questionnaire

HOOS Hip Dysfunction and Osteoarthritis Score

HOOS The Hip disability and Osteoarthritis Outcome Score

HREC Human Research Ethics Committees

HREC Human Research Ethics Committee

IA intra-articular

ICC intraclass correlation coefficient

IHLC Internal Health Locus of control

IRT Item response theory

JSW joint space width

LOA limits of agreement,

KLG Kellgren and Lawrence grade

KOOS Knee Injury and Osteoarthritis Score

KOOS Knee injury and Osteoarthritis Outcome Score

m metres

MCID minimal clinically important difference

MCII minimal clinically important improvement

MDT multi-disciplinary team

MFES modified fall efficacy scale,

MHLC Multidimensional Health Locus of Control

MIC minimal important change

MID minimal important difference

mJSW minimum Joint Space Width

MNSQ Infit & outfit mean square statistics,

MnSq mean square

MRI magnetic resonance imaging

MSK musculoskeletal

N/A not applicable

N/A Not Applicable

NNFI Non-normed Fit Index,

NRS Numeric Rating Scale

NS non-significant,

NSAIDs non-steroidal anti-inflammatory drugs

NSW New South Wales

OA Osteoarthritis

OACCP Osteoarthritis Chronic Care Program

OAMP Osteoarthritis Management Programs

OARSI Osteoarthritis Research Society International

OMERACT-OARSI Outcome Measures in Rheumatology- Osteoarthritis Research Society International

OR odds ratio

OSES osteoporosis self-efficacy scale,

p p-value

PAM-13 Patient Activation measure

PAM-13 Patient Activation Measure 13 item version

PAM-MH Patient Activation Measure Mental Health version

PBC Perceived behavioural control

PCA Principal Components Analysis

PDP Perceived difficulty for physicians

PDPh Perceived difficulty for pharmacists

PEPPI-10 Perceived Efficacy in Patient–Physician Interactions ten item

PEPPI-5 Perceived Efficacy in Patient–Physician Interactions five item

PHLC Powerful Others Health Locus of control

PRISMA Preferred Reporting Items for Systematic Reviews and Meta-Analyses

PROs patient-reported outcomes

PsA Psoriatic arthritis

r correlation coefficient

RA Rheumatoid arthritis

RCT randomised controlled trial

RM Rasch model

RM Rasch model

RMSEA root mean square error of approximation,

SB $\chi 2$ Satorra-Bentler chi-squared statistic,

SCQOA The Stages of Change Questionnaire in Osteoarthritis

SD Standard Deviation

SEE-C self-efficacy for exercise scale.

SEP Self-efficacy for physicians

SEPh Self-efficacy for pharmacists

SF-36 Short form 36

6MWT Six Minute Walk Test

SLE Systemic Lupus Erythematosus

SPSS Statistical Package for the Social Sciences

SRMR standardized root mean square residual,

SS Systemic sclerosis

3-D 3 dimensional

TJA total joint arthroplasty

TUG Timed up and go

UK United Kingdom

US ultrasound

VAS visual analogue scale

WOMAC Western Ontario and McMaster Universities Osteoarthritis Index

WOMACG Western Ontario and McMaster Universities Arthritis Index Global score

Y year

YLD years lived with a disability

β Beta coefficient

χ2 Chi squared

χ2MH Mantel-Haenszel chi-squared statistic

CHAPTER ONE

Chapter One: Thesis introduction

1.1. The burden of osteoarthritis

Symptomatic osteoarthritis (OA) is often defined by the presence of pain and/or stiffness in a joint, coupled with structural changes seen on imaging. The experience of pain and/or stiffness often prompts people to seek treatment for their OA. The nature, severity and impact of osteoarthritis symptoms vary greatly between individuals and change over time (1). Osteoarthritis is a highly prevalent condition (2, 3). The knee, hip and hand joints are the most commonly affected sites of OA (4). OA prevalence is influenced by multiple factors including age, sex, cultural background and type of employment. The prevalence of OA has been reported to be between 7% and 16% for lower limb OA (5, 6) and 7 and 15% for hand OA (7).

Osteoarthritis is not only highly prevalent, it is a severely disabling condition. Of 291 conditions included in the global burden of disease study, hip and knee OA together ranked as the 11th highest contributor to global disability (8). This is particularly worrisome given that in the presence of an aging population, and increasing rates of obesity, the prevalence of OA and its projected burden, are projected to increase dramatically (3). The financial burden of OA is also substantial to both individuals and the economy. Furthermore, these costs increase with disease severity (9). The overall impact of OA also needs to be considered within the context of multimorbidity; OA is one of the most common comorbidities of people already living with chronic conditions such as heart disease, diabetes or cancer (10). Not only

is the health management of people with complex multimorbidity more difficult, the ensuing health outcomes of these people may be worse (11).

1.2. Non-surgical interventions for hip and knee osteoarthritis

In order to address the needs of those suffering with this chronic, disabling condition, international management guidelines have been developed to recommend evidence-based care for OA. These guidelines are in broad agreement regarding their recommendations of efficacious non-pharmacological and pharmacological treatments for management of hip and knee OA (12). There is relative consensus amongst these guidelines that hip and knee OA management should be tailored to the individual and include the following three core effective, non-surgical, non-pharmacological interventions.

i) Self-management and OA education: Education and support for OA self-management are consistently considered cornerstones of OA treatment (12, 13), despite providing modest treatment effects when used in isolation (14, 15).

ii) Exercise: Land-based exercise has been found to provide moderate treatment effects in the short term for both pain and function (16) in knee OA, and slightly smaller, short-term treatment effects have been found in hip OA (17).

iii) Weight loss interventions: Weight loss is consistently recommended as a core intervention in management guidelines for people with hip or knee OA who are overweight or obese (12, 13). In people with knee OA who are obese, weight loss has been demonstrated to provide moderate treatment effects for pain and function. These outcomes improve further when the weight loss is greater than 5% of baseline bodyweight (18).

In addition, OA management guidelines recommend that these three core, non-pharmacological OA interventions may be supported by judicious and regularly monitored use of pharmacological agents (13). Other evidence-based adjunctive treatments also recommended, as required, include : psychological interventions such as pain coping skills or management of psychological symptoms (19); referral to a physiotherapist if the patient is weak or stiff; walking aids and other assistive devices (20); braces, taping and specialised footwear (21) in the presence of joint malalignment (13).

1.3. Osteoarthritis management programs

Despite the existence of clear, evidence-based recommendations for OA management, their implementation into clinical practice has been limited (22). Considerable discrepancy persists between the recommendations and the actual care that is received by OA patients in Australia. The Caretrack study estimated that only 43% of Australians with OA received evidence-based care (23), which is consistent with international evidence (24, 25). The reasons for this evidence-practice gap are complex. One contributing factor is that there remains a perception amongst healthcare providers that OA is merely a 'normal part of the ageing process'. This perception is exacerbated when patients present with OA within the context of competing demands imposed by multimorbidity, particularly in the presence of perceived 'more serious' conditions such as cardiovascular disease and cancer (26). When OA care is provided it is often limited to pharmacological treatments alone, and non-pharmacological treatments such as exercise and weight loss are underutilised or overlooked

entirely (25). In response to these challenges, models of OA care have been developed and implemented over the past 15 years to attempt to close this evidence-practice gap (22).

Implemented models of OA care can be referred to as OA management programs (OAMPs). As there is no published definition of what constitutes an OAMP, for the purposes of this Thesis, an OAMP is defined as a model of evidence-based, non-surgical OA care that has been implemented in a real-world setting providing the following four components:

a. Individualised OA care
b. Provided as a package of care with longitudinal reassessment and treatment progression
c. Two or more components of the core, non-surgical, non-pharmacological interventions (i.e. self-management and OA education; exercise; weight loss interventions)
d. Optional evidence-based adjunctive treatments as required.

Published international OAMPs differ markedly in their models of service delivery and implementation across a range of healthcare systems (22). However OAMPs include the core components previously described in common of i) education and support for self-management, ii) exercise; and generally offer various combinations of other evidence-based therapies such as: weight loss interventions; psychological support; review of analgesics; assistive devices and braces (22).

1.3.1. Outcomes of osteoarthritis management programs

There is some evidence in previous literature to support the efficacy or value of OAMPs in terms of improvements in pain and functional ability (27-30). The focus of this Thesis,

however, was not on efficacy, but on the relationships between clinical characteristics and outcomes following participation in an OAMP. If clinical characteristics were found to be associated with OAMP outcomes, this knowledge could be used to direct services to people likely to improve following participation. In an era where the delivery of evidence-based care is demanded within an environment of growing economic rationalism, the ability to predict outcomes to intervention would enable identification of people with OA most likely to benefit from an OAMP and improve resource allocation. Further, people identified as unlikely to improve, or likely to worsen despite treatment, may be targeted for adjunctive therapies that may improve their chances of success e.g. cognitive behavioural therapy or motivational counselling. Therefore, the overarching research question of this Thesis was:
'Can clinical presentation predict response to a non-surgical chronic disease management program for hip and knee osteoarthritis?'.

1.3.2. Responders and non-responders

The **first aim** of this Thesis was to determine if *"baseline clinical characteristics were associated with the outcomes of an OAMP".* One method of reporting outcomes following OA interventions is to present the proportion of people who achieve a pre-defined threshold for change in the outcome(s) of interest. Participants who achieve this threshold of improvement in an outcome can be classified as 'responders' and those who did not as 'non-responders'. There are few previous studies that have reported outcomes following OAMPs in terms of the

proportion of participants who were 'responders', and these used varying definitions of 'response' (27, 29, 30) . One study that adopted this approach was a randomised controlled trial that compared a patient self-management program- "the Enabling Self-Management and Coping of Arthritic Knee Pain Through Exercise (ESCAPE-knee pain) program" with usual primary care in the United Kingdom (27). A pre-defined threshold to classify participants as responders was set at 15% of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) function score (27). Sixty-one percent of participants were classified as responders immediately following the intervention and 54% of all participants maintained this at the six-month follow up (27). Furthermore, two cohort studies reported outcomes from a Dutch stepped-care model for hip and knee OA. The first study defined the responder threshold using the Outcome Measures in Rheumatology- Osteoarthritis Research Society International (OMERACT-OARSI) response criteria (31). The authors reported that 47% of participants were responders at 12-weeks (30). The second study used the threshold of pain Numeric Rating Scale (NRS) ≤4 to indicate the threshold for responders to the intervention (29). Twenty-nine percent of participants with complete data were responders according to this threshold (29).

The rationale for the definition of responder for our cohort studies is discussed in detail in the Thesis methods in **Chapter Two**. In brief, we used the threshold of treatment response described by Angst et al (2002) that was developed in a prospective cohort study of participants with hip and knee OA engaged in a multimodal rehabilitation intervention (32). This threshold included: a relative change greater or equal to 18% (100 × (change of

score/baseline score); and an absolute change of 9 points improvement of WOMAC global scores at 26-week assessment compared to baseline. We classified cohort participants as responders and non-responders in **Chapter Four** of this Thesis.

1.3.3. Worsening outcomes following OAMPs

In addition to the concept of responders and non-responders, it is important to consider that people can report worsening outcomes despite their participation in OAMPs. There is little evidence available in the current literature that addresses worsening outcomes following OAMPs. Further, there is a paucity of evidence available to support the use of a suitable threshold of worsening. In **Chapter Five** of this Thesis, we compared three different thresholds of worsening following participation in an OAMP and examined the relationships of these thresholds with baseline clinical characteristics. The thresholds of worsening and the rationale for the choice of these thresholds is also described in **Chapter Two.**

1.3.4. Predictors of 'response' and 'worsening' following OAMPs

Few strong predictors of response to OAMPs have been identified in previous studies. Four previous studies, which attempted to identify baseline predictors of response to three discrete OA management programs, did not find consistent predictors of response (29, 30, 33, 34). Furthermore, we were unable to find any studies in the literature concerned with the clinical characteristics associated with worsening outcomes following an OAMP. The study in **Chapter Five** was therefore undertaken to investigate characteristics that were associated with

worsening outcomes. It compared three different thresholds of worsening, following participation in an OAMP, and examined the relationship of worsening with baseline clinical characteristics.

1.4. Osteoarthritis self-management attitudes and capabilities

Self-management is a general term used to describe an individual's ability to manage the physical and psychological symptoms, treatments, consequences and lifestyle changes required to live with their condition (35). As stated earlier in section 1.3 of this chapter, a core intervention of OAMPs is to support participants to self-manage their OA. It follows that OA self-management attitudes and capabilities may be associated with the ensuing clinical outcomes following participation in an OAMP. We hypothesized that better attitudes towards and/or capabilities regarding self-management of OA at baseline would be associated with improved outcomes following participation in an OAMP.

To test this hypothesis, it was necessary to identify a suitable instrument to measure attitudes towards and/or capabilities regarding self-management of OA. Measures of this construct have not been widely used to report OAMP outcomes, and there is no 'gold standard'. The **second aim** of this Thesis was therefore to *"identify a suitable standardised instrument that could be used to measure OA self-management attitudes and/or capabilities."* To make a recommendation on the "best" measure of OA self-management attitudes and capabilities, it was necessary to consider the measurement property or psychometric evidence available for

instruments measuring the construct in the population of interest. Measurement properties provide information on the performance of an instrument such as construct and structural validity, reliability, sensitivity, responsiveness and feasibility within a given setting (36, 37). **Chapter Six** comprises a systematic review that synthesized the available measurement property evidence for instruments assessing OA self-management attitudes and capabilities.

The **third aim** of this Thesis was *"to test the measurement properties of an instrument measuring OA self-management attitudes and capabilities in people living with OA."* The Patient Activation Measure (PAM-13), identified in the systematic review, was selected as a potential instrument for further investigation. The measurement properties of the PAM-13 were tested in a cross-sectional study described in detail in C**hapter Seven.**

The **fourth aim** of this Thesis was *"to determine if OA self-management attitudes and capabilities predict response to an OAMP"*. We tested the hypothesis that that better attitudes towards and/or capabilities regarding self-management of OA, measured on the PAM-13, at baseline, would be associated with improved outcomes following participation in an OAMP in **Chapter Eight**. In this longitudinal cohort study, we examined the relationship between patient activation (PAM-13 scores) and changes in pain and function.

1.5. Aims of this Thesis

To recap, the overarching research question of this Thesis was:

'Can clinical presentation predict response to a non-surgical chronic disease management program for hip and knee osteoarthritis?'.

There were four specific aims of this Thesis that addressed this question:

I. To determine if baseline clinical characteristics were associated with outcomes of the OACCP (**Chapters Three, Four and Five**)

II. To identify a suitable standardised instrument to measure OA self-management attitudes and/or capabilities (**Chapter Six**)

III. To test the measurement properties of an instrument measuring OA self-management attitudes and capabilities in people living with OA (**Chapter Seven**)

IV. To determine if OA self-management attitudes and capabilities predict response to the OACCP (**Chapter Eight**)

1.6. Thesis outline

The Thesis is arranged in nine chapters, each written so that they can be read independently. The University of Sydney permits published manuscripts arising from the candidature to be included in the Thesis. Six chapters of this Thesis consist of papers that were submitted for publication. Four of these chapters have been published with the remaining two currently under review. Each of these six chapters addresses one of the four specific aims.

Following on from the Thesis introduction comprising **Chapter One,** the methods used in this Thesis are described in **Chapter Two**. This chapter provides a broad overview of data collection for the four cohort studies and statistical techniques used to model potential predictors of outcomes following an OA management program (OAMP). There is an emphasis on less commonly-used methodologies including the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) initiative methodology for systematic reviews of outcome measurement instruments and Rasch statistical modelling to test the measurement properties of outcome measurement instruments.

The ***first aim*** of this Thesis was to identify baseline characteristics that were associated with "response" to an OAMP. To address this aim we first conducted a literature review to determine if clinical characteristics had been shown to be associated with response to OAMPs in previous work. The body of literature addressing predictors of response to OAMPs was very limited. To broaden the depth of our knowledge of the literature, we conducted a narrative review to summarise existing evidence of predictors of response to any nonsurgical interventions (pharmacological and non-pharmacological) for OA. This review is presented as **Chapter Three** and also considered whether the authors defined a threshold for 'response' within the analyses. It is presented as published in the *Rheumatic Clinics of North America* (38).

The study presented in **Chapter Four** was a prospective, longitudinal, clinical cohort study that examined the relationships between baseline characteristics of OA management program participants and *positive response* to the intervention. We hypothesised that *"it would be possible to predict participants likely to "respond" to the program using baseline demographic, psychological, disease-related and functional performance variables"* (39). This study is presented as published in *The Journal of Rheumatology (39).*

The second prospective, longitudinal, clinical cohort study presented in **Chapter Five** investigated participant characteristics associated with *worsening* outcomes following participation in an OA management program. We hypothesised that: *"the same participants with similar demographic, psychological, disease-related and functional performance predictor variables would be identified as 'worse' across three definitions of worsening"*. It is presented as published in *Arthritis Care and Research (40).*

The ***second aim*** of this Thesis was to identify a suitable standardised instrument to measure OA self-management attitudes and/or capabilities. This aim was addressed with the systematic review in **Chapter Six** presented as published in *Osteoarthritis and Cartilage (41).* The aims of the systematic review were to:

"i) identify studies reporting measurement properties of instruments assessing attitudes toward and/or capabilities regarding self-management of OA;

ii) systematically critique the studies evaluating instruments using the Consensus-based Standards for the selection of Health Measurement Instruments (COSMIN) tool: and iii) synthesize the evidence available with the possibility of making rudimentary recommendations concerning the best evidence-based instruments…"(41).

One instrument identified from the systematic review, deemed worthy of further investigation, was the Patient Activation Measure 13 (PAM-13). The PAM-13 measures self-reported knowledge, skill, and confidence for self-management of one's health (42). Although there was limited extant measurement property evidence for this instrument, it was chosen because it appeared to have adequate face validity to justify further testing in an OA population. The ***third aim*** of this Thesis was to test the measurement properties of an instrument measuring OA self-management attitudes and capabilities in people living with OA. **Chapter Seven** presents the cross-sectional cohort study undertaken to examine the measurement properties of the PAM-13 presented in this Thesis as submitted to *BMC Health Quality of Life Outcomes (under review)*.

After adequate measurement properties were established for the PAM-13 in **Chapter Seven**, the ***fourth aim*** of this Thesis was to determine if OA self-management attitudes and capabilities were associated with changes in pain and function following participation in an OAMP. **Chapter Eight** describes a longitudinal cohort study. Finally, **Chapter Nine** provides

a summary of the principal findings of this Thesis, discusses the implications of these findings and proposes directions for future research.

1.7. Ethical approvals

Ethical approval was obtained for **Chapters Four and Five** from the NSW Population and Health Services Research Ethics Committee AUREI Reference HREC/12/CIPHS/63. Cancer Institute NSW Reference Number 2012/08/413.

Ethical approval for **Chapters Six and Seven** was provided by Human Research Ethics Committees (HREC):

Hunters Hill Hospital NSPHEC 2016-LNR-007; Mount Wilga Hospital NSPHEC 2017-LNR-005 and Northern Sydney Local Health District reference: RESP/16/11, HREC reference: LNRl16/HAWKE/14.

The remaining chapters did not require ethical approval.

CHAPTER TWO

Chapter Two: Thesis Methods

Detailed methodology regarding each of the studies contained within this Thesis are included within the discrete chapters reporting them. The purpose of this chapter is to provide brief overview of the main methodological considerations of this PhD work including:

2.1 The Osteoarthritis Chronic Care Program- the OAMP that is the main focus of this Thesis;

2.2 Definition of response;

2.3 Definition of worsening;

2.4 Regression models: logistic and linear multivariable regression;

2.5 COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) initiative methodology for systematic reviews of outcome measurement instruments and;

2.6 Rasch modelling.

2.1. The Osteoarthritis Chronic Care Program Model of Care

The Osteoarthritis Chronic Care Program (OACCP) was developed and implemented by the Musculoskeletal Network of the Agency for Clinical Innovation (ACI); a pillar of the New South Wales (NSW) Ministry of Health, Australia. The OACCP model of care was based on the chronic care model (CCM) (43), a proactive healthcare delivery system that enables patients to self-manage their condition. This CCM was supported by a coordinated team of health care providers with appropriate clinical expertise, all underpinned by suitable health information technology systems (44). Using CCM principles the OACCP aimed to support

participants with hip and knee OA to reduce their pain, increase their function and quality of life through provision of a combination of evidence-based interventions delivered by a multidisciplinary team.

In this model of care, the OACCP Musculoskeletal (MSK) coordinator was an experienced Physiotherapist or Exercise Physiologist with expertise in the management of OA and the central health professional who coordinated the care of OACCP participants. At the first program visit the MSK coordinator conducted a detailed assessment, provided participants with education about their condition and supported participants to set goals for the management of their OA and any comorbidities. The assessment and goal setting was based on the HealthChange Australia Model of Health Change™ (45). This model addressed important aspects of behaviour facilitating long-term behaviour change including; building motivation; identifying and addressing barriers and facilitators to health behaviours; and establishing the patient-practitioner relationship (46). This model also informed behaviour-change strategies developed in partnership with participants.

Participants were also prescribed an individualised exercise program incorporating both strength and cardiovascular training, which was progressed at scheduled face-to-face reassessments at 12-, 26-, and 52-weeks. Following their first OACCP visit, participants attended a multidisciplinary clinic for consultation with a rheumatologist and a selection of other health care professionals according to their individual clinical needs. Some examples of multidisciplinary clinic consultation included:

- a dietitian review to manage weight loss in participants who were overweight/obese
- an occupational therapy review for those requiring assistive devices
- a social work review for participants requiring psychosocial support
- an orthotist review for those presenting with knee joint malalignment to consider bracing interventions.

The effects of these management strategies were measured using standardised clinical outcomes at baseline, 12, 26 and 52 weeks of the OACCP (47). This enabled evaluation of clinical outcomes at a patient level, and of OACCP outcomes at group level.

The OACCP model of care has been implemented at various public teaching hospitals in NSW and is considered a state-of-the-art program for OA management. There are also programs at two private metropolitan hospitals. These programs are based on the OACCP model of care but are known as the Osteoarthritis Management Program (OAMP). The OACCP study sites included in this Thesis were Royal North Shore Hospital, and Wollongong Hospital (both public facilities) whilst the OAMP study sites were Hunters Hill Private and Mount Wilga Private Hospitals. Given that all sites implemented the OACCP model of care, all public and private sites are referred to generally as the OACCP throughout this Thesis.

2.2. Definition of 'response'

One of the important methodological considerations of this Thesis was to define the threshold used to indicate that a participant was a 'responder' in the longitudinal cohort study described in **Chapter Four**. Three thresholds of response identified in previous literature were considered

(1) The OMERACT-OARSI response criteria. This required either:

"*Improvement in pain or in function ≥ 50% and absolute change ≥ 20 points, or;*

Improvement in at least two of the following:

- *Pain ≥ 20% absolute change ≥ 10 points*
- *Function ≥ 20% absolute change ≥ 10 points*
- *Patients' global assessment ≥ 20% absolute change ≥10 points"* (31).

(2) The minimal clinically important improvement (MCII) proposed by Tubach et al (2005). This was globally defined as "the smallest change in measurement that signifies an important improvement in a patient's symptom" (48). More specifically the MCII was defined as follows:

"*For knee and hip OA, MCII for absolute (and relative) changes were, respectively, (a) -19.9 mm (-40.8%) and -15.3 mm (-32.0%) for pain on a visual analogue scale; (b) -18.3 mm (–39.0%) and -15.2 mm (-32.6%) for patient's global assessment on a visual analogue scale; (c) -9.1 (-26.0%) and -7.9 (-21.1%) for Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) function subscale score."* (48).

(3) The threshold proposed by Angst et al. (2002). This was based on the concept of the minimal clinically important difference (MCID), which can be can be defined as the smallest difference in scores of the variable concerned that is considered beneficial by participants of the intervention (49). This MCID required a relative change greater or equal to 18% (100 × (change of score/baseline score) and an absolute change of 9 points improvement of WOMAC global scores at 26-week assessment compared to baseline (32).

Of the three thresholds considered above, both the OMERACT-OARSI response criteria and the thresholds proposed by Tubach et al. (2005) were relatively stringent, requiring large improvements in outcomes to indicate a response. These thresholds were proposed to determine the efficacy of pharmacological therapies in randomised controlled trials. In these contexts, it was necessary to set large thresholds to account for the larger treatment effects required to justify the risks of side-effects associated with drug therapies (31, 48). The study described in **Chapter Four** of this Thesis did not require such a large threshold of improvement to indicate a positive response as the OACCP interventions were relatively safe, and their efficacy had already been tested. Hence, we chose the third threshold of response proposed by Angst et al. (2002). This threshold was developed in a prospective cohort study of participants with hip and knee OA engaged in a multimodal rehabilitation intervention (32). It was based on a similar intervention to the OACCP, with comparable expected treatment effects and a lower risk profile (compared with OA drug trials). Further details of this definition and how it was applied are presented in **Chapter Four**.

2.3. Definition of worsening

The aim of the study in **Chapter Five** was to investigate baseline characteristics that predicted the worsening of symptoms associated with the treated joint experienced by participants of the OACCP. There were relatively few studies that described thresholds used to indicate symptomatic worsening of the treated joint. Given the uncertainty regarding an appropriate threshold, we decided to compare three different definitions to test if predictors of worsening were similar across the definitions:

(1) Worsening based on the study by Angst et al. (2002). This MCID for worsening was 9.6 points absolute and 21% relative-change in the WOMAC Global score. This threshold was derived from the same study used for the *threshold for response* described in **Chapter Four** (32), which investigated a comprehensive rehabilitation intervention for hip and knee OA.

(2) Worsening based on a transition question proposed by Jordan et al. 2009. The transition question was proposed as a suitable indicator of worsening following criticism of the use of pre/post patient-reported outcomes (PROs), such as the WOMAC, in the measurement of OA self-management education program effectiveness (50). Interview data was considered to represent the gold standard outcome of this study, however, the responses to the PROs were found to be a poor reflection of the interview data (50). It was proposed that this dissonance was possibly due to 'response shift' whereby a change in the participants' perspective occurred following engagement in such programs (50). Further, a single transition question was

found to be more reflective of the interview data (50). The transition question required participants to report the global magnitude and direction of change experienced regarding the treatment joint using a 5 point Likert scale (5=much worse; 1=much better) following the intervention (50). Participants were asked at OACCP reassessments "Compared with when I started this program, my hip/knee has…": "much improved", "moderately improved" "slightly improved", "not changed", "slightly worse", "moderately worse", or "much worse". In the absence of evidence for an ideal cut-off for change on the transition question that is meaningful to participants, the threshold "moderately worse" was chosen in an attempt to ensure that participants were reporting a change that was important to them (51).

(3) Worsening based on a composite outcome. This outcome included either or both WOMAC Global score MCID and transition question definitions of worsening. This composite outcome was chosen to reflect clinical practice whereby concerns would be raised if a patient declared their joint was feeling moderately worse following treatment or if considerably worse PRO scores were obtained. The composite outcome reflected the dual importance of the two clinical outcomes with the added benefit that it increased the power of the analysis by combining two outcomes of common aetiology (52).

2.4. Regression models

2.4.1. Multivariable logistic regression

Logistic regression modelling was used to test the relationships between baseline characteristics and dichotomous response outcomes in **Chapters Four and Five**. To make sure that the models were adequately powered, we ensured that there were at least ten predictors per event. This also avoided overfitting the model with too many predictors (53). For both studies, the first modelling step examined the association between each of the predictor variables and their response using univariable logistic regression analyses.

The methods used to determine the base multivariable model differed in **Chapters Four and Five.** The multivariable modelling in **Chapter Four** used a forced entry technique. This involved all the variables being entered into the base model. In contrast, the procedure to determine the base model in **Chapter Five** involved considering the results of the univariable analyses that were conducted for each of the three definitions of worsening. To enable interpretation and comparison of results between the composite definition and the single-outcome definitions of worsening, variables exhibiting odds with $p < 0.2$ that trended in the same direction across a minimum of two definitions were included in the base models for all three definitions of worsening (52).

After base models had been established, the analyses proceeded similarly for both studies. In an iterative manner, the variable associated with the largest P-value was removed from the

base model until only variables with $P \leq 0.05$ remained. To control for confounding, the regression coefficients of the remaining variables were checked on removal of each variable from the model and in the presence of a change of 10% or more the variable was retained. Possible interactions were tested by combining variables, and then testing the relationship between the combined variable and the outcome variables. The Hosmer-Lemeshow goodness-of-fit test was used to test how well the chosen models fit the data; a p-value of >0.05 indicated the model fit was adequate (54). Further specific details of how the modelling steps were applied are described in **Chapters Four and Five**.

2.4.2. Multivariable linear regression

In the study comprising **Chapter Eight**, linear regression modelling was used to examine relationships between baseline PAM-13 scores and changes in WOMAC pain and function scores of participants with knee OA following 12 and 26 weeks of the OACCP. The decision to use linear regression modelling was partially a pragmatic one. The sample size required to analyse data using logistic regression is larger than what is required to run adequately powered linear regression analyses (53-55). The variables used in the analyses were normally distributed, an important assumption of linear regression, making it possible to examine the size and direction of the effect patient activation, along with other variables of interest, on changes in pain and function. Only data from people who chose the knee joint as their treatment joint was used in the analyses. This decision was made because there were four times the number of people with knee OA in the cohort compared to those with hip OA.

Hence, the smaller mixed sample for this study was unrepresentative of people with hip OA. This could have led to difficulties in generalising the results of the study to people with hip OA.

The first step of linear regression modelling was unadjusted analysis with simple linear regression to study relationships between baseline variables and WOMAC Pain and Function change scores. In the second step, models included an adjustment for baseline WOMAC Pain/Function scores to test whether the baseline score was associated with the covariates and the outcome scores at 12- and 26 weeks. Variables with a P-value of ≤ 0.1 were included as candidate variables in multivariable models which were built using a backward selection, forced entry technique. We used the more conservative threshold of P ≤ 0.1 because this study had a much smaller sample size, we wanted to avoid overfitting the regression model with too many variables and hence maintain adequate statistical power (55). In a similar manner to the techniques used in **Chapter Four**, at each step, the least significant variable (p-value<0.05) was removed from the base model and the beta coefficients were checked. If the coefficients changed by ≥ 10% the variable was retained in the model as a confounder, if not it was removed. Further details of these analyses are documented in **Chapter Eight**.

2.5. The COSMIN systematic review method

The COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) initiative method for performing systematic reviews of measurement property

studies (56) was used to guide the systematic review in **Chapter Six** of this Thesis. The COSMIN initiative aims to: "improve the quality of studies on measurement properties by developing methodology and practical tools for assessing measurement properties" (57) .

Briefly, a taxonomy of measurement properties was provided by the COSMIN initiative (58) and a search filter provided to capture all studies concerned with these measurement properties (59). The COSMIN initiative has devised a method to determine the methodological quality of measurement property studies (60) and developed a checklist of criteria for good measurement properties to rate the measurement property results for each instrument (61). The quality ratings of the study methodology and the quality of the measurement property results for each instrument were combined to provide a best evidence synthesis. A more detailed explanation of the COSMIN initiative methodology for systematic reviews used in this Thesis is provided in **Chapter Six**.

2.6. Rasch Modelling

Chapter Seven examined the measurement properties of the PAM-13 in a cohort of OA management program participants. The PAM-13 was developed by Hibbard et al. (2005) using Rasch modelling (42). A Rasch analysis was used to test the fit of our data to the expected model and provide measurement property evidence for the PAM-13 in a sample of people living with OA. The Rasch model (RM) is a specific item response theory measurement model that is used to develop instruments and is capable of scaling raw

observed scores (measuring a latent trait) into linear variables. A "latent trait" refers to a construct that cannot be observed directly (e.g. the mathematics ability of a child) (62). The RM provides a mathematical framework against which the fit of data from another sample can be compared to quantify how well it fits the test (62).

The RM assumes that responses to the test items are affected by the ability of the person and the difficulty of the item (62). A RM describes the probability of success (in the case of PAM-13 that an item will be 'agreed' with) based on the difference between the ability of the person (person ability) and the difficulty of the item to achieve (item difficulty) (62). 'Person ability' is calculated using the number of items of the instrument that a person agreed with. 'Item difficulty' is estimated using the number of persons in the sample who agreed with an item.

Essential requirements of the RM include adequate data-model fit and confirmation of unidimensionality. Measures of fit, namely infit, outfit and standardised fit, are used to confirm that the instrument conforms to RM requirements and to indicate how accurately or predictably data fit the model (63). Unidimensionality refers to one dimension or attribute of the subject being measured at a time. This is a requirement of the RM to enable items of an instrument to become more difficult with each subsequent item (62). A particular type of Principal Components Analysis (PCA), unique to RM, is used to provide evidence of unidimensionality (63). The study comprising **Chapter Seven** used the partial credit model to examine model-data fit, tested the assumption of unidimensionality and interpreted Rasch

analysis results for the PAM-13. More detailed description of the methods used are described in that Chapter.

CHAPTER THREE

This chapter includes the following published literature review:

Eyles J, Lucas B, Hunter DJ. Targeting Care Tailoring Nonsurgical Management According to Clinical Presentation. Rheumatic Disease Clinics of North America. 2013; 39: 213- 33.

Chapter 3

‘Targeting care: tailoring non-surgical management according to clinical presentation.’

Authors

Jillian Eyles[1,], Barbara R Lucas[1], David J Hunter[2,3,4]

Affiliations: [1]Physiotherapy Department, Royal North Shore Hospital; [2]Department of Rheumatology, Royal North Shore Hospital; Northern Clinical School; [3]Kolling Institute of Medical Research; [4]Institute of Bone and Joint Research; University of Sydney;

Corresponding author: Jillian Eyles

Statement from co-authors confirming authorship contribution of the PhD candidate

The co-authors of the paper: "Targeting care: tailoring non-surgical management according to clinical presentation", confirm that Jillian Eyles has made the following contributions:

- Conception and design of the research
- Analysis and interpretation of the findings
- Writing of the manuscript and critical appraisal of the content

In addition to the statements above, in cases where I am not the corresponding author of a published item, permission to include the published material has been granted by the corresponding author.

Jillian Eyles Date: 1st March, 2019

As supervisor for the candidature upon which this Thesis is based, I can confirm that the authorship attribution statements above are correct.

David Hunter Date: 1st March, 2019

Chapter Three: Targeting care: tailoring non-surgical management according to clinical presentation

3.1. Abstract

This review summarizes the evidence available for patient characteristics that have been analysed as potential predictors of response to non-surgical interventions for patients with hip and knee OA. The specific variables targeted for this review include: body mass index, psychological factors, muscle strength, tibiofemoral alignment, radiographic changes and signs of inflammation. Several studies provide moderate to good evidence of potential predictors of response to non-surgical treatments, while areas for future research are illuminated.

3.2. Introduction

The activity limitation attributed to osteoarthritis (OA) places it within the world's top 10 most disabling conditions (64). Globally OA affects approximately 18.0% of women and 9.6% of men over 60 years of age (65). In 2003 the annual costs of OA and other rheumatic conditions was an estimated $128 billion to the United States economy (66). These enormous costs are projected to rise steeply with the steadily increasing prevalence of rheumatic conditions (66).

This prevalent, expensive, disabling disease is incurable, so it follows that current treatments focus on symptomatic relief. Commonly reported treatment goals for this group include reductions in joint pain, stiffness, activity limitation, participation restriction, and improvements in quality of life and well-being. To assist clinicians in achieving these goals with their patients, numerous international evidence-based guidelines for management of hip and knee OA have become available (67-72). There is uniformity in most of the recommendations made by the guidelines (73) and agreement that conservative management of hip and knee OA should combine both non- pharmacological and pharmacological treatment modalities (67-72).

The recommendations made in the guidelines for management of hip and knee OA are broad. For example: the evidence-based, expert consensus guidelines from the Osteoarthritis Research Society International (OARSI) (2008) include no fewer than 20 recommendations for the non-surgical management of hip and knee OA (74), however, the treatments are not

arranged systematically to indicate the order of priority in which they should be undertaken. With so many recommended management options tabled, it would be advantageous to know which treatments are most likely to be effective for the individual with hip or knee OA according to clinical presentation.

This review examines the evidence available for identification of clinical characteristics which predict patient response to nonsurgical treatments for hip and knee OA. The summation of this evidence may assist clinicians to target treatments most likely to benefit patients according to clinical presentation, and identify areas for further research

3.3. Body Mass Index

Obesity is a known risk factor for the development of arthritis (75) and is a strong predictor for long-term progression of the disease (76). There is evidence that obesity is a risk factor for knee OA, however, the relationship between obesity and the risk of developing hip OA is less clear (77, 78). International guidelines nonetheless recommend weight reduction in people with hip and knee OA who are overweight or obese (68, 71, 73, 79). There is strong evidence that weight loss is an effective treatment for knee OA, yet little evidence exists regarding weight loss as an effective treatment for obese patients with hip OA.

It seems reasonable that body mass index (BMI) may be a clinical characteristic that is predictive of response to weight loss interventions, surprisingly evidence exists that it does not. A post hoc analysis of a randomised controlled trial (RCT) involving 111 overweight

veterans with knee OA, investigated nine clinical characteristics as possible predictors of weight change between baseline, 16 and 32 weeks. The minimum amount of weight loss required to define treatment responder was not provided. Multi-regression analysis revealed that BMI was not predictive of weight loss in response to the interventions for overweight veterans with knee OA (80). The external validity of this study was limited by confining recruitment of participants to veterans.

Two studies found that BMI was not predictive of response to a Dutch multi-modal, stepped care model of pain management for hip and knee OA. Snijders and colleagues (2011) investigated the efficacy of the Dutch model in a cohort of 183 participants with hip and knee OA (30). The model combined pharmacological and non-pharmacological treatments. Two possible definitions of positive treatment response were described: (i) Outcome Measures in Rheumatoid Arthritis Clinical Trials/ Osteoarthritis Research Society International (OMERACT-OARSI) Responder Criteria and (ii) patient reported numeric rating scale (NRS) for pain ≤ 4. At 12-week reassessment, 86 patients were responders according to definition (i), and 71 fulfilled definition (ii). BMI was one of eleven potential predictors of response included in analyses, and was not a significant predictor of response to this program (30), however, the study was underpowered to identify true predictors of response. A more recent study utilized the same Dutch model, focussing specifically on a stepped-care protocol used to progress the use of acetaminophen and non-steroidal anti-inflammatory drugs (NSAIDs) at standardised intervals according to patient-reported pain levels (29). The definition of treatment responder was patient-reported NRS pain ≤ 4, 100 participants met

this target. The study was underpowered to analyse 13 patient characteristics, including BMI, as possible predictors of response. Further research is required to determine whether BMI is predictive of positive treatment response achieved by participants of this multi-modal stepped-care model of pain management.

Two well-powered studies were identified examining the potential of BMI as a predictor of response to Cyco-Oxygenase-2 (COX-2) inhibitors. Bingham et al. (2011) pooled the results of two similar RCT's comparing the efficacy of Etoricoxib and Celecoxib to placebo (81). The OMERACT-OARSI Responder Criteria determined that 562 participants were responders to the COX-2 inhibitors following 12 weeks of the intervention. BMI was one of 16 variables analysed as potential predictors of response, BMI failed to predict positive treatment response to the COX-2 inhibitors (81). Similar results were found by Detora et al. (2001) combining the results of three 6 week RCT's comparing the COX-2 inhibitor Rofecoxib with placebo in 1501 patients with hip and knee OA (82). Responder criteria were not defined. Patient data was analysed according to subgroups representing 14 baseline characteristics including BMI. Analysis of covariance failed to identify any baseline measures associated with treatment response (82). To date, good evidence exists that baseline BMI does not predict the response of patients with hip and knee OA treated with COX-2 inhibitors.

A single study explored BMI as a predictor of response to intra-articular corticosteroid (CSI) for management of hip OA. Robinson et al. (2007) followed 120 patients with hip OA for 12 weeks following CSI (83). Participants were classified as responders to the CSI at 12 weeks if

>15% reduction in baseline WOMAC pain sub-scale was achieved, 48 participants met this criterion. Logistic regression determined that BMI, one of 14 variables analysed, was not a significant predictor of response to hip CSI (83) however the study was underpowered to detect true predictors of treatment response.

Four cohort studies were identified exploring possible predictors of response to intra-articular (IA) hyaluronic acid derivatives for hip and knee OA. Short-term efficacy and tolerability of IA Hylan G-F 20 were assessed in 4253 patients with symptomatic knee OA (84). Responder criteria were not defined, the primary outcome was pain measured at baseline and 3 weeks on a 4-point Likert scale. At three-weeks post IA Hylan G-F20, 88.4% of patients assessed their pain as better or much better. Logistic regression of 7 potential predictors of short-term pain reduction determined that underweight patients were more likely to report reduced knee pain than their obese counterparts. The method of recruitment threated the validity of this evidence; the authors invited 840 orthopaedic surgeons to who report on at least five consecutive patients receiving Hylan G-F 20 for relief of knee OA pain, introducing significant selection bias.

Longer term outcomes of patients with knee OA receiving IA Hylan G-F were explored in the following three cohort studies (85-87). A retrospective cohort of 155 patients with knee OA was reassessed 7-14 months following IA Hylan G-F 20 (85). The definition of responder was not specified. Analysis of 16 possible predictors found that BMI was not a significant predictor of patient satisfaction, though this study was underpowered to identify the possible

predictors, and the retrospective design was prone to significant recall bias. Longer term outcomes of Hylan G-F 20 were also studied in a small cohort of 32 patients with mild to moderate knee OA 6 months following IA Hylan G-F 20 (86). Clinical response was defined using the OMERACT-OARSI "high improvement" criteria, only 15 participants were responders, and 8 variables, including BMI, were investigated as predictors of response, leaving the study underpowered to detect significant predictors. BMI was not significantly correlated to patient response. A prospective cohort study examining 84 patients with knee OA for 6 months following knee IA Hylan G-F20 found that Short Form -36 health survey scores for were significantly improved at 6 months post injections (87). Responder criteria were not described. Three factors including the subjects' percentage above ideal body weight were analysed for correlations with positive treatment outcomes seen on the Short Form-36 health survey: Physical Function, Role-Physical and Role-Emotional categories. The subjects' percentage above ideal body weight was not predictive of improvements. The high number of patients lost to follow up (23%) affected the validity of this study. Evidence for BMI and % above ideal bodyweight as clinical characteristics predictive of longer-term response to knee IA Hylan G-F20 was inconclusive due to low validity and power.

BMI was not a significant predictor of response of people with hip OA to IA Hylan G-F20. Migliore et al. (2008) evaluated 250 patients with hip OA who received intra-articular Hylan G-F20 (88). Treatment response was defined as a ≥30% improvement in baseline Lequesne scores or NSAID usage at 6 months, the number of participants classified as responders was unclear. Ten possible predictors of treatment response were analysed, BMI was not a

significant predictor of response to hip IA Hylan G-F 20 (88). The large number of drop-outs (42%) affected the validity of this study.

Patients with a lower BMI may be more likely to experience a reduction in chronic knee pain following treatment with glucosamine sulphate. A prospective correlational study of 39 participants with chronic knee pain followed patients receiving 1.5g daily glucosamine sulphate for 12 weeks (89). Participants were not required to have been diagnosed specifically with OA, affecting the external validity of this study. The definition of treatment responder was not described and 7 patient characteristics were examined to as potential predictors of reduction of pain rated on a VAS. The study was underpowered to determine the effects of 7 potential predictors.

To date, most of the evidence suggests that BMI is not a consistent predictor of response to non-surgical treatments for people with hip and knee OA. Some evidence exists that BMI is not predictive of response to a weight loss program in overweight veterans with knee OA (80). There is good evidence that BMI does not predict response to COX-2 inhibitors for hip and knee OA management (81, 82). The evidence is weak that BMI is not predictive of treatment response to either a multi-modal stepped-care pain management model (29, 30), hip CSI (83), hip IA Hylan G-F20 (88), or glucosamine for chronic knee pain (89). The evidence for BMI as a predictor of response to knee IA Hylan G-F is weak and conflicting (84-87). Further research is required to determine whether BMI is a clinical characteristic that can foretell response to non-surgical treatments for people with hip and knee OA.

3.4. Psychological factors

Complex interactions exist between psychological factors and perceived symptoms of OA. Compared with their peers, people with OA report increased prevalence of depression and depressed mood (90). The intensity of perceived OA pain has been demonstrated to be predictive of depression severity in this cohort (90). Poor mental health has been associated with worse overall hip and knee OA pain and deterioration in mental health has been found to precede short-term exacerbations of OA pain (91). Treatment of depression in people with arthritis appears to improve depressive symptoms, reduce OA pain, improve function and quality of life (92) and therefore is an important consideration in the management of OA.

Many treatments prescribed for hip and knee OA management, particularly exercise, weight loss programs and medications; require active participation from the patient. The compliance with and efficacy of these treatments may be influenced by the individual's mental state to affect rehabilitation outcomes. The prospective cohort study "Predictors for response to rehabilitation in patients with hip or knee OA" (34) featured 250 patients with hip and knee OA who participated in a 3- 4 week multimodal rehabilitation program combining exercise therapy, hydrotherapy, relaxation strategies, distraction techniques, patient education, manual therapy, thermotherapy, and electrotherapy. Participants were assessed at baseline and 6 months following the program. Three different definitions of treatment responder were used: (i) the minimal clinically important difference (MCID) (18%) improvement shown on the WOMAC, (ii) improvement on the Transition scale, and (iii) MCID improvement on WOMAC and improvement on the Transition scale. The transition scale was described as a

measure of the current state of health of the OA joint compared with its state 6 months earlier (34). There were 21 personal, lifestyle and psychological measures investigated as potential predictors of the three definitions of responder. Depression and anxiety were evaluated using the Hospital Anxiety and Depression Scale, mental health was assessed using the mental component of the SF-36. The absence of depressive symptoms was determined to be a strong predictor of all three of the responder definitions suggesting that depression may hinder the achievement of positive treatment outcomes of patients with hip and knee OA following a 3-4- week rehabilitation program. This study did not attempt to answer the question as to why they did not achieve the same results as their non-depressed counterparts, but we could hypothesize that perhaps those patients with depression have more difficulty complying with a comprehensive rehabilitation program. This is an interesting area for further research (34).

The presence of depression may affect the ability of overweight people with OA to lose weight. A post hoc analysis of an RCT aimed to identify predictors of positive treatment response resulting from weight loss interventions for 111 overweight veterans with knee OA (80). Veterans were randomized into groups receiving 24 weeks of nutritional counselling, a home exercise program, a combination of both or usual care. There were no differences in weight loss between intervention groups and 9 variables were investigated as possible predictors of weight change between baseline, 16 and 32 weeks of the RCT. The amount of weight loss required to indicate successful treatment response was not indicated. Symptoms of depression were evaluated using The Centre for Epidemiologic Studies Depression Scale which measured 20 items to achieve a score out of 60. The presence of depression was

indicated by a score ≥ 16. The absence of depression was the only independent predictor of weight loss at 16 weeks and 32 weeks (80). This study did not define treatment responder; however, it did suggest that depressive symptoms may limit the ability of veterans to lose weight.

Depression and anxiety did not seem to predict treatment response of patients with knee OA to CSI. A small study of 59 patients with knee OA receiving CSI examined 10 possible predictors of a favourable response, defined as a ≥ 15% reduction of pain rated on VAS, to injection of methylprednisolone acetate (93). The Hospital Anxiety and Depression score at baseline was not found to consistently predict treatment response. Given that 59 patients were used to investigate 10 predictors of response, this study was underpowered to detect meaningful effects of the potential predictors.

Mental health scores did not seem to predict response to a combined non- pharmacological and pharmacological pain management program. Predictors for response to analgesics were explored in relation to a cohort study of 347 patients investigating treatment outcomes of a stepped model of care for hip and knee OA. The model initially offered education, lifestyle and weight loss advice, physiotherapy, acetaminophen, then progressed to other medications at intervals as guided by a pain numerical rating score (29). Treatment response was defined as achievement of pain NRS ≥4, there were 100 responders. Thirteen possible predictors of response were explored including mental health. The Short Form-36 (SF-36) questionnaire was used to assess health-related quality of life and the mental component summary (MCS)

scores of the SF-36 were used to reflect mental health. Mental health rated by the MCS was not a significant predictor of response to the stepped model of pharmacological pain management for patients with hip and knee OA, however, this study was underpowered to analyse 13 possible predictors.

Self-reported participant mood failed to predict treatment outcome in a small cross- over RCT of 11 patients with osteoarthritis receiving two different NSAIDs (94). During two treatment periods of four weeks duration participants received ketoprofen and piroxicam. A 4- week wash-out period followed the initial drug treatment prior to commencement of the second drug. Participants were classified as treatment responders if they showed ≥30% improvement of 5 of the 7 variables measured at baseline including; pain, tenderness, swelling, patient and physician global assessments, acute-phase protein levels and disability. There were 20 baseline variables explored as possible predictors of response including mood, assessed using an 18-item questionnaire. Mood was not a significant predictor of treatment outcome; however, the small sample size of this study leaves it underpowered to detect meaningful effects of the predictors investigated.

In summary, two well-designed, adequately powered studies used specific measures of depression that were predictive of response to intervention (34, 80). Both studies demonstrated the relationship between the absence of depressive symptoms and positive non-pharmacologic treatment outcomes. The treatments investigated in these studies included a comprehensive rehabilitation program and weight loss interventions. These treatment

modalities require high levels of active participation of the patients involved, which may be affected by the presence of depressive symptoms. Interestingly, the three studies investigating drug therapy regimes, perhaps not requiring such a high level of active participation by the subjects, consistently found different measures of psychological factors incapable of predicting treatment response (29, 93, 94). Two of these studies were underpowered (93, 94), and the third that was inadequately powered did not measure depression specifically.

3.5. Muscle Strength

In view of the biomechanical influence and protective functions of skeletal muscles surrounding joints, muscle weakness is considered to be an important possible factor in the development and progression of OA. Evidence for the significance of muscle strength in the pathogenesis of OA remains unclear (95, 96). Higher quadriceps strength may have a protective effect against the development of symptomatic OA (96). Whether muscle weakness precedes the onset of OA, or if it is a feature of already established disease seen on X-ray, or is only related to the onset of pain and other symptoms, is an area for further research.

The evidence for the role of muscle strength in the progression of OA is varied. Limited evidence exists to support muscle strength as a predictor of knee OA progression (76). Yet, over time, people with knee OA who have greater quadriceps strength report less pain and superior functional ability compared to their weaker counterparts (97). Quadriceps strength has been studied widely in relation to knee OA, however muscles around the hip stabilising the pelvis also have an effect on adduction forces around the knee that may result in increased

compression of the medial compartment (98) and influence the pathogenesis and progression of OA. Hip OA has also been associated with significantly reduced lower limb muscle strength (99), however limited evidence is available to explain the role of hip and thigh musculature in the development and progression of the disease. Further research is required to explain this possible relationship.

Treatments for hip and knee OA have long included specific exercises designed to strengthen muscles surrounding the joints involved. High-level evidence exists regarding the reduction of pain and dysfunction in knee OA through therapeutic exercises (100). The evidence for the efficacy of exercise in hip OA to date is less convincing (101) yet exercise is often prescribed. Wright et al. (2011) published a study aiming to identify baseline characteristics of patients with hip OA likely to respond favourably to Physical Therapy interventions (102). As part of a larger RCT, 91 patients were randomised to groups receiving manual therapy, exercise therapy, a combination of both or usual care. The OMERACT-OARSI responder criteria determined treatment responders. Ten variables were analysed as predictors of treatment response. Measures of muscle strength using a hand-held dynamometer were not predictive of treatment success. Only 22 of the 68 participants were responders which left the study underpowered to identify predictors of response (102).

There has been recent interest in the nature of lower limb muscle weakness in people with knee OA. Decreased quadriceps strength in knee OA has been attributed to both loss of muscle cross-sectional area (103) and reduced ability to activate the muscles (104). In a

cohort of 111 subjects taken from a larger RCT, baseline ability to activate quadriceps was examined as one of 9 possible predictors of changes in strength of the muscle following a 6 - week exercise program for subjects with knee OA. Primary outcome measures were quadriceps strength and quadriceps activation, measured using a burst-superimposition maximum isometric quadriceps torque test, however a definition of treatment response was not identified. Although lower quadriceps activation was associated with lower strength, the baseline quadriceps activation did not predict the magnitude of gain in quadriceps strength following exercise therapy (105). These results suggest that patients with OA should benefit from strengthening exercises regardless of baseline quadriceps activation.

Baseline muscle strength does not seem to predict the degree of symptomatic relief achieved following a weight loss program in obese people with knee OA. The 192 participants, who were part of larger RCT, were randomised to two different dietary interventions. A significant response to the interventions included the OMERACT-OARSI responder criteria and improvement on Knee Injury and Osteoarthritis scores (KOOS). Although weight loss was achieved in most of the subjects, only 64% achieved the OMERACT- OARSI responder criterion. There were 23 variables investigated as possible predictors of response to the weight loss programs, including measurements of baseline hamstrings and quadriceps strength using isometric dynamometry. Baseline muscle strength was not predictive of symptomatic relief in response to the weight loss program (106). The study was underpowered to detect significant predictors from a possible 23 variables.

One study investigated muscle strength as a predictor of response to a pharmacological agent. Jones et al. (1996) performed a cross-over RCT comparing CSI with saline (placebo) in 59 subjects with knee OA (93). Ten possible predictors of treatment response were analysed. Treatment response defined as ≥ 15% decrease in pain rated on a visual analogue scale, was not predicted by baseline quadriceps strength measured using a commercial strain gauge. This study was significantly underpowered to analyse 10 predictors of response.

Further research into the role of muscle strength in the pathogenesis of OA and subsequent progression of the disease may be helpful in refining recommendations for therapies aimed at OA prevention and further joint deterioration as a consequence of OA. To date, muscle strength has not been demonstrated to predict response to non-surgical interventions for hip and knee OA

3.6. Tibiofemoral Joint Alignment

Varus (bow-legged) or valgus (knock-kneed) tibiofemoral joint alignments are clinical characteristics observed in some people with knee OA. Joint alignment impacts the distribution of load borne by the medial and lateral compartments of the articular surface of the knee. Static knee alignment is conventionally determined using full- length weight-bearing radiographs of the lower limb with knees extended. Lines are drawn from the centre of the femoral head to the talus through the middle of the femoral and tibial shafts to indicate the loadbearing mechanical axis, then measurements made of various angles subtended from where those lines intersect (107-109). Neutral alignment is commonly defined as 0-2 degrees

of varus (110) meaning that in a normal knee the mechanical axis passes medial to the knee joint resulting in 60-70% of weight-bearing forces to pass through the medial articular surface (111). Varus malalignment results in higher loads borne through the medial compartment of the knee, whereas increased compressive forces through the lateral articular surface accompany valgus malalignment.

Dynamic knee alignment can be assessed using three dimensional (3-D) gait analysis. In varus knees, the measurement of knee adduction moment during the stance phase of walking is an indirect measure of joint compressive forces sustained within the medial tibiofemoral joint compartment (109, 112, 113). Static and dynamic alignment are important to consider in view that altered distribution of forces placed through the joint surface may lead to damage of articular structures, possibly increasing the risk of OA development or worsening existent disease.

It remains unclear whether knee joint malalignment precedes incident knee OA (107, 108), however varus alignment considered to be a significant predictor of knee OA disease progression (76). Knee malalignment has been demonstrated to interact with other risk factors for OA progression, increasing the likelihood of disease acceleration. Possible interactive factors include greater quadriceps strength (114), the stage of disease observed in the individual (115) and obesity (108).

The evidence for the relationship between knee malalignment and reported OA symptoms remains unclear (115, 116). Nevertheless, some non-surgical treatments in OA management guidelines aim to reduce pain and dysfunction associated with tibiofemoral malalignment. Orthotic bracing, shoe wedges and muscle strengthening are recommended with a view to improving the biomechanics of the joint (69-72). Several studies have investigated knee joint alignment as a predictor response to non-surgical management of OA. An RCT by Lim et al. (2008) examined the effect of a 12-week quadriceps strengthening program on knee adduction moment, pain, and function in 107 subjects with knee OA. Knee alignment was assessed on radiographs and participants stratified according to whether they had more neutral (< 5 degrees) or more varus (≥5 degrees) alignment. Specific responder criteria were not described. Patients in the strengthening group achieved significant improvements in strength regardless of alignment. Self-reported function, performance measures and knee adduction moment determined using 3-D gait analysis were unchanged by the intervention in both alignment groups. Pain, assessed using the WOMAC pain subscale, was significantly improved in the strengthening group subset that was more neutrally aligned. Neutral knee joint alignment may mediate improvements in knee OA pain following a 12-week quadriceps strengthening program (117).

Immediate changes in static alignment and knee adduction moment were not predictive of response to lateral wedge insoles at 3 months. A cohort of forty volunteers with knee OA were provided with laterally wedged insoles to assess the immediate effects of the insoles on knee OA pain, knee adduction moment and static alignment (109). The lateral wedges immediately

reduced knee adduction moment calculated using 3-D gait analysis and walking pain measured using the WOMAC pain subscale, but had no effect on static alignment as determined on full-length leg radiographs. Alignment was defined as the angle subtended by the intersection of the femoral and tibial mechanical axes. Varus malalignment was determined when the angle was <180 degrees with valgus indicated by >180 degrees. After 3 months of wearing the insoles, significant improvements in pain and function persisted. A definition of treatment responder was not utilised; 10 predictors at baseline of outcome to intervention at 3 months were explored. Neither immediate changes in static alignment or knee adduction moment were predictive of decreased pain and improvement in function 3 months following the intervention (109). The size of this cohort limited the ability of this study to identify true predictors of response to the intervention.

A larger RCT of 192 obese subjects with knee OA allocated patients to two different weight loss interventions (106). Knee joint alignment was assessed using a 'Plug-in Gait model' with a six-camera stereophotogrammetric system and markers on anatomic landmarks. A knee was categorised as varus when the alignment was > 0 degrees, and valgus if < 0 degrees. Baseline knee alignment was one of 23 variables examined as possible predictors, however, it failed to predict improvements in Knee Injury and Osteoarthritis Outcome Scores (KOOS) or achievement of OMERACT- OARSI Responder Criterion following weight loss interventions (106). In view that only 64% of patients were treatment responders according to OMERACT-OARSI criterion, the study was underpowered to detect effects of significant predictors.

It is interesting to consider the definitions of knee malalignment utilised in the three studies discussed above. Lim et al. (2007) employed a more extreme definition of ≥5 degrees to indicate varus malalignment (117). In contrast to the two other studies they categorized subjects to knee malalignment groups if the mechanical axis did not appear as a straight line (106, 109). This may have increased the severity of malalignment observed within the participants assigned to the varus group investigated by Lim and colleagues, compared to the subjects categorized to knee malalignment groups in the other studies. Participants with varus malalignment studied by Lim et al. (2007) did not experience improvements in pain following strength-training, while the neutrally-aligned reported significant pain reduction (117). Perhaps the higher severity of varus malalignment was key to the determination of knee joint alignment as a predictor of outcome to intervention in this study. Future research considering knee malalignment as a predictor of treatment response to conservative treatments should consider carefully the definition of joint alignment.

3.7. Radiographic and MRI Assessment

The presence of radiographic osteophytes (OP) and joint space narrowing are commonly used to diagnose OA. These features are combined to determine radiographic disease severity according to scoring systems such as the Kellgren and Lawrence grade (KLG) (118). Despite known limitations, radiographs are inexpensive, accessible and easy to interpret, so are commonly used in research for classification of subjects to determine eligibility and for stratification of samples according to radiographic severity. Radiographic joint space width

(JSW) or minimum joint space width (mJSW) is recommended for use in clinical trials, however, MRI is preferred particularly for assessment of cartilage morphology (119).

Relatively few papers analyse radiographic severity of hip and knee OA as possible predictors of response to non-surgical, non-pharmacological treatments. Two of the three papers identified doing so examined the ability of radiological and MRI OA severity to predict response to weight loss interventions. A small RCT of 30 obese female participants with knee OA compared two dietary weight loss interventions (120). Within the intervention group, 90% of participants achieved clinically significant weight reduction of >10%, and 33% had a 50% improvement in knee OA symptoms. A strict definition of treatment responder was not provided. Structural joint damage was assessed at baseline using both the KLG classification, and low field MRI (0.2T) to assess various measures of cartilage abnormalities, bone marrow lesions (BML's) effusions and synovitis of the medial, lateral and patellofemoral compartments of the knee. Five baseline radiographic characteristics and clinical outcomes following the weight loss interventions were investigated for correlations. None of the imaging variables were able to forecast symptomatic response to treatment (120) however this study was likely underpowered to identify significant predictors.

A second RCT randomized 192 obese patients with knee OA into 8 weeks of two experimental dietary interventions (106). Results were calculated for the entire cohort as the method of weight loss was not relevant for this analysis. OA symptoms were evaluated at baseline and 16 weeks using the OMERACT- OARSI Responder Criteria and changes in

KOOS. High field MRI was assessed using the Boston-Leeds Osteoarthritis of the Knee Score (BLOKS) to measure joint damage at baseline. Conventional radiography determined the baseline KLG and minimum Joint Space Width (mJSW). MRI and radiographic measures failed to find any relationship between variables assessing knee structural damage and symptomatic improvements following the dietary interventions (106). Only 64% of patients were treatment responders according to OMERACT-OARSI criterion, therefore this study may also be insufficiently powered to detect effects of 23 potential predictors.

A third study examining the ability of radiographic features to predict response to non-surgical, non-pharmacological interventions was conducted by Hinman et al (2008) (109). A cohort of 40 patients with knee OA wore full length 5- degree lateral wedge insoles for 3 months. Improvements were observed in WOMAC pain and function subscales following the intervention. Tibiofemoral OA severity was assessed at baseline using the KLG scoring system. Following analysis of 10 possible predictors of outcome, greater disease severity indicated by higher KLG scores were predictive of worse pain at 3 months. This study did not define responder criteria, and the small sample size reduced the ability to identify predictors of response to lateral wedge insoles.

Two studies examine radiographic severity using KLG as potentially predictive of response to interventions combining both non-pharmacological and drug therapies for hip and knee OA. Both investigated cohorts of patients with hip and knee OA participating in a Dutch multi-modal, stepped-care pain management program (29, 30). During the 12-week program

subjects received standardised non-pharmacological management and pain relieving medications prescribed and altered at set intervals depending on self-reported pain at reassessment. The definition of positive treatment response in the initial cohort of 183 patients was fulfilment of either the OMERACT-OARSI Responder Criteria or NRS ≤ 4 (30). The later study of 347 subjects required NRS ≤ 4 at 12 weeks to indicate successful response to the intervention (29). Both studies analysed OA severity as determined by KLG scores as possible predictors of positive treatment outcomes. The first study tested 11 possible predictors of response to intervention and found that disease severity did not forecast improvements in overall pain and function as a result of the 12-week pain management program. (30). In the second study, 13 predictors were tested for correlation with treatment response at the 4 different steps of the treatment model. Greater OA severity was independently associated with a higher chance of pain relief achieved in response to use of acetaminophen (29). This correlation was discovered because unlike the first study, the predictors of response were tested at each of the separate steps of the program. There were 59 responders to acetaminophen (paracetamol), so the study was underpowered to test 13 predictors. Although the evidence is tenuous, this finding lends support to the recommendations made by international OA management guidelines to trial acetaminophen as a first line pharmacological treatment of hip and knee OA (67-72), even in those patients with severe disease.

Evidence to the contrary was presented by Case et al. (2003) in the results from a double-blind, placebo-controlled RCT comparing the efficacy of acetaminophen and diclofenac

sodium for pain management of knee OA (121). Eight-two patients were randomized to 3 groups receiving either one of the drugs or placebo. The primary outcome at baseline, 2 and 12 weeks was the WOMAC scale. The diclofenac sodium group alone achieved significant improvement (≥20%) in all 3 WOMAC subscales following the intervention. The subjects were stratified according to pre-study medication, baseline pain and disease severity indicated by KLG, in order to identify subsets of patients that were consistent in response to the treatments. None of the sub-groups consistently demonstrated a preferential response to acetaminophen or diclofenac sodium. This study suffered from a high number of dropouts (>25%). Three of the five subjects who withdrew from the diclofenac sodium treatment arm (n=25) did so as a result of adverse effects. Despite the evidence presented in this paper for the superior efficacy of diclofenac sodium, the relatively high risk of unwanted side effects lends further weight to the OA treatment guidelines recommending a trial of acetaminophen prior to commencing NSAID therapy (67-72) and it can be presumed that this follows regardless of radiographic severity.

Four articles were identified investigating radiological predictors of response to CSI for hip OA. Of these, only one reported that radiographically determined disease severity was a significant predictor of positive response to steroid injection (122). This retrospective cohort study reviewed radiographs, radiology reports and medical records of 361 patients who had received fluoroscopically guided IA 80 mg depomedrol or methylprednisolone with bupivacaine. The definition of treatment responder was a 50% decrease in pain reported on a visual analogue scale (VAS). Immediate positive response to injection was evident in 68.2% of

hips and delayed response were apparent in 71.4%. OA severity was measured at baseline using KLG classification and the grades were split into groups for analysis. Multivariate regression determined that radiographic severity of OA was an independent predictor of treatment response. Patients with advanced disease were much more likely to experience both immediate and delayed onset of pain relief. The authors suggested that people with advanced hip OA are likely to achieve better response to CSI than those with mild or moderate disease (122). Although a good number of participants were recruited, these inferences should be considered cautiously in view of the inherent risk of bias associated with the retrospective cohort design of this study.

In contrast, Robinson et al. (2007) utilizing a similar fluoroscopically guided injection of methylprednisolone and bupivacaine into the hip joint of 120 people with hip OA concluded that radiographically determined OA severity was not predictive of response to intervention (83). This cohort study assessed symptomatic response to 40 mg and 80mg dosages of the steroid. A decrease in the WOMAC pain by >15% was considered to indicate positive treatment response, 75 patients were classified as responders at 6 weeks. The authors concluded that the higher dose (80mg) of methylprednisolone was more effective and lasted longer. Twelve possible predictors of treatment response included KLG scoring. Forward logistic regression found that KLG was incapable of predicting reduced pain in response to hip CSI (83). This study was underpowered to detect predictors of response among 12 variables.

Similar conclusions were made regarding a small prospective cohort of 27 patients with hip OA assessed at baseline, 2, 12 and 26 weeks following hip IA lignocaine and methylprednisolone (123). The main outcome measure was pain measured on VAS. The degree of radiological severity according to KLG classification and mJSW had no significant bearing on the reported pain relief following hip steroid injection, however, the small sample size decreased the power to detect significant predictors. The fourth RCT compared ultrasound-guided CSI to IA hyaluronic acid, saline (control), and standard care (no injection) in 77 subjects with hip OA (124). Response to treatment was delineated by the OMERACT-OARSI Responder Criteria, there were 14 responders to steroid injection. CSI was significantly more effective than the 3 other treatments. Univariate regression analysis determined that of 5 predictors analysed, radiographic severity using Croft grading and mJSW were not predictive of treatment response to CSI, however, the study was underpowered to analyse 5 predictors (124). Further research is required to explore the value of radiographic and MRI clinical characteristics indicating disease severity as potential predictors of response to hip CSI.

Three cohort studies attempted to identify radiographic characteristics of patients with hip and knee OA that were predictive of treatment response to IA Hylan G-F 20(85, 86, 88). Migliore et al. (2008) followed 250 patients who received US-guided IA hylan G-F 20 into OA hips (88). Treatment response was defined as an improvement of ≥30% Lequesne index or NSAID use. Significant improvements were reported for all outcome measures at 3,6, 9 and 12 months when compared to baseline. Multivariable regression analysis of 8 baseline

variables determined that KLG was unable to predict treatment response (88). A high number of dropouts limited the validity of this study. The second study followed a small cohort of 32 patients with mild to moderate knee OA for 6 months following knee IA Hylan G-F 20 (86). The OMERACT-OARSI "High improvement" responder criteria for OA were utilized to define responders to treatment. Fifteen participants met the responder criteria. Eight predictors of treatment response were explored, including mJSW which was not predictive of positive response to Hylan G-F 20 injection (86). The study was underpowered and limited by the exclusion of patients with severe OA. Conrozier et al. (2003) (85) studied a cohort of 155 patients across the spectrum of mild through to severe knee OA. Knee joint space loss in a single compartment seen on radiograph and meniscal calcinosis noted on MRI scans were predictive of a good outcome to knee IA Hylan GF-20. The definition of treatment responder was not utilised in this retrospective cohort, (85). Weak evidence exists that knee joint space loss in a single compartment and meniscal calcinosis may predict response to IA Hylan G-F 20, and this warrants further research.

Bennett et al. (2007) (89) investigated the symptomatic response of 39 subjects with chronic knee pain treated with 1.5g oral Glucosamine sulphate for 12 weeks. Primary outcome measures at baseline and 3 months included pain VAS rated on movement, VAS for restriction in function and patient-rated global change score. These outcomes were all found to be significantly improved at 12 weeks but the responder criteria were not specified. Seven possible predictors of reduced pain and improved functional ability were analysed using regression modelling. The authors concluded that lower levels of PFJ osteophytes, BMI and

functional self-efficacy, were predictors of successful glucosamine treatment. The presence of osteophytes within the medial and lateral compartments of the tibiofemoral joint was not correlated with response to the intervention (89).The study was underpowered to identify true predictors of response and participants did not require formal diagnosis of OA, so these results must be viewed accordingly.

Overall, there was weak evidence that radiographic measures of OA severity may have predictive value in the identification of potential responders to lateral wedged insoles (109), CSI (122) and glucosamine sulphate. MRI assessment was predictive of response in a single study concerned with knee Hylan G-F 20 injections (85). A greater number of studies exist that were unable to predict response to treatment based on radiographic disease severity or MRI. There is good evidence that KLG scores are not predictive of response to hip CSI (83). Further research is required to clarify the roles radiography and MRI perform as clinical characteristics that predict response to non-surgical treatment for hip and knee OA.

3.8. Inflammation

Abnormal progressive remodelling of joint tissues occurs in response to local inflammatory processes arising within osteoarthritic joints (125). Physical examination may reveal clinical signs such as presence of joint swelling, effusion and heat. With recent improvements in imaging techniques, synovial hypertrophy has become a surrogate marker of local inflammation within a joint.

Signs of inflammation were examined as potential predictors of response to weight loss interventions for participants with knee OA. The clinical cohort described by Gudbergsen et al. (2012) participated in a 4 month dietary intervention (106). Responders were required to fulfil the OMERACT-OARSI responder criteria. Joint damage severity was assessed on MRI using the BLOKS which included scoring for synovitis and effusion. Although synovitis and effusion were not predictive of OMERACT-OARSI response, there was some evidence that the effusion score correlated with changes in the KOOS ADL score from baseline to 4 months. Responder criteria for the KOOS score were not provided. There were 23 variables assessed as potential predictors of response, 123 patients were responders, therefore this study was insufficiently powered for this many predictors. The presence of inflammatory markers such as effusion and synovitis requires further investigation as predictors of symptomatic response to weight loss interventions for overweight patients with OA.

Systemic pharmacological agents such as NSAID's and COX-2 inhibitors are prescribed for their analgesic properties and also to reduce inflammatory activity in affected joints. Two studies found that the presence of swelling was not predictive of response to these drug therapies. The data of three 6 week RCT's comparing rofecoxib with placebo were combined to analyse the consistency of response of patients with hip or knee OA classified into subgroups determined by 14 demographic and disease factors (82). Three outcome measures were analysed in relation to the subgroups; Pain walking on flat surface (WOMAC), Patient Global Assessment of Response to Therapy, and Global Assessment of Disease Status. The definition of treatment responder was not provided. Overall, the subgroups did not show

consistent interactive effects with all three outcome measures. The absence of knee swelling of participants with knee OA significantly correlated with improved scores on the Patient Global Assessment of Response to Therapy, but not the 2 remaining outcome measures (82). Another study investigated swelling among numerous possible predictors of response of patients with OA and rheumatoid arthritis to ketoprofen and piroxicam (94). The trial was very small with 11 participants with OA, so was underpowered to determine significant predictors of response (94). Further investigation into signs of inflammation as possible predictors of response to NSAID's and COX-2 inhibitors would be helpful to the clinician attempting to tailor pharmacological management according to clinical presentation.

Intra-articular corticosteroids aim to directly reduce inflammatory processes occurring within joint tissues. An RCT by Chao et al. (2010) examined inflammatory characteristics assessed on ultrasound (US) as predictors of response to intra-articular corticosteroid injection for knee OA (126). Participants were categorised as 'inflammatory' if synovial hypertrophy (synovitis) with or without effusion was detected on grayscale US examination of the affected knee(s) at baseline. Within the intervention group 16 patients presented with synovitis on US, and 18 did not. At 4 weeks there were no significant differences between the inflammatory and non-inflammatory subgroups. Significantly lower WOMAC pain scores of the non-inflammatory sub-group at 12 weeks suggested that those without inflammatory characteristics experienced prolonged beneficial effects from corticosteroids. The presence of effusion had no influence on response to corticosteroid injection (126).

The presence of hip joint synovitis on US assessment of patients with hip OA was predictive of treatment response to CSI. An RCT compared standard care (no injection), injection of normal saline (placebo), non-animal stabilised hyaluronic acid and methylprednisolone acetate (124). Of the participants receiving CSI, 14 participants were classified as responders according to the OMERACT-OARSI criteria. The authors concluded that synovitis was predictive of response at 4 and 8 weeks following injection, however, this study was underpowered to establish clear associations between the 5 variables analysed as possible predictors. In contrast, Robinson et al (2007) found that evidence of hip synovitis and effusion on US were not predictive of clinical response to intra-articular methylprednisolone injection (83). The cohort study defined response to intervention as >15% reduction in baseline WOMAC pain score at 6 and 12 weeks following injection. This study was also underpowered to identify predictors of response. Further research using greater numbers of subjects is required to explore US determined inflammatory characteristics as predictors of response to CSI of osteoarthritic hips.

Inflammatory characteristics identified on physical examination of patients with knee OA failed to predict response to CSI. The presence of local inflammation indicated by knee joint fluid, local heat, synovial thickening and stiffness were explored as possible predictors of response to intra-articular methylprednisolone in an RCT of 59 participants with symptomatic knee OA (93). No predictors of response were identified, perhaps a result of this study being underpowered. Pendleton et al. (2008) examined similar clinical signs of inflammation; presence of heat, effusion and synovial thickening in addition to the presence

of effusion and synovitis on knee US, as predictors of improvements in baseline WOMAC pain scale 1 and 6 weeks following CSI (127). The presence of heat was associated with 29% greater reduction in night pain, otherwise clinical and US inflammatory signs were not predictive of response. The study was underpowered and did not publish any measures of data variability. Adequately powered well-designed research is necessary to determine whether clinical and US signs of inflammation are predictive of outcomes following CSI for knee OA.

Moderate effusion was associated with good outcome following intra-articular injection with Hylan G-F 20 in patients with symptomatic OA. Conrozier et al (2003) followed a cohort of 155 patients who received three intra-articular hylan G-F 20 injections and were evaluated 7-14 months later (85). Treatment outcomes included patient satisfaction, safety, changes in pain and function which were assessed on 4-point Likert scales. This study was limited by the lack of validated outcome measures and the retrospective study design.

Only one study investigated signs of inflammation as predictive of outcome to non-surgical, non-pharmacological intervention. There is weak evidence that that synovitis and effusion seen on MRI are unable to predict response to weight loss in patients with OA (106). Numerous studies were concerned with signs of inflammation as predictors of outcomes to pharmacological agents, but few were sufficiently powered. Some evidence exists that knee joint swelling may predict good outcomes from rofecoxib (82) and that patients without synovitis observed on US experience prolonged pain relief following CSI injection compared

to those patients with knee OA presenting with synovitis (126). The evidence for synovitis on US as a predictor of response for outcomes following hip CSI is conflicting (83, 124). There is little evidence to support the use of clinical inflammatory signs as predictors of response to CSI for knee OA (93, 127) . Further research is required to determine whether signs of inflammation are useful predictors of response to conservative therapies for people with hip and knee OA.

3.9. Other clinical characteristics that may predict response to intervention

For the purposes of this review, we selected patient characteristics that we deemed interesting to examine as potential predictors of outcome to interventions for people with hip and knee OA. There is a wider range of presenting features than those covered here, analysed as potential predictors of response and further discussed in the literature. Among the articles identified through literature searches for our chosen predictors, age and gender were commonly analysed as potential predictors of response to intervention, but appeared to hold little predictive capacity overall. Four well-powered studies investigating predictors of response to; COX-2 inhibitors (20, 21), a rehabilitation program (32), and exercise therapy (45) provided moderate to good evidence that age was not a powerful predictor of response to these interventions. Further investigation of age as a predictor of response to alternative interventions for patients with hip and knee OA is justified.

One adequately powered study determined female gender to be a characteristic predictive of treatment success following participation in a rehabilitation program (34). In contrast, three

well-powered studies found gender not predictive of treatment success for COX-2 inhibitors and exercise therapy (81, 82, 105). Additional research into gender as a predictor of treatment response to different non-surgical modalities is required.

Pain and function WOMAC subscales are often used as primary outcome measures in OA research. Of the studies extracted from literature searches performed around our chosen predictors, two well-designed studies examined WOMAC pain and function scores as predictors of response to COX-2 inhibitors. One found that that baseline WOMAC pain was not predictive of response to Etoricoxib and Celecoxib. Lower levels of function on the WOMAC decreased the odds of response to the drugs, but the difference in WOMAC function scores between responders and non-responders was not clinically significant (81). The second study concluded that baseline WOMAC function was not predictive of response to Rofecoxib (82). Although baseline WOMAC pain and function scores were not predictive of response to COX-2 inhibitors, these measures may prove to be interesting predictors of response to different non-surgical interventions in other research.

3.10. Summary

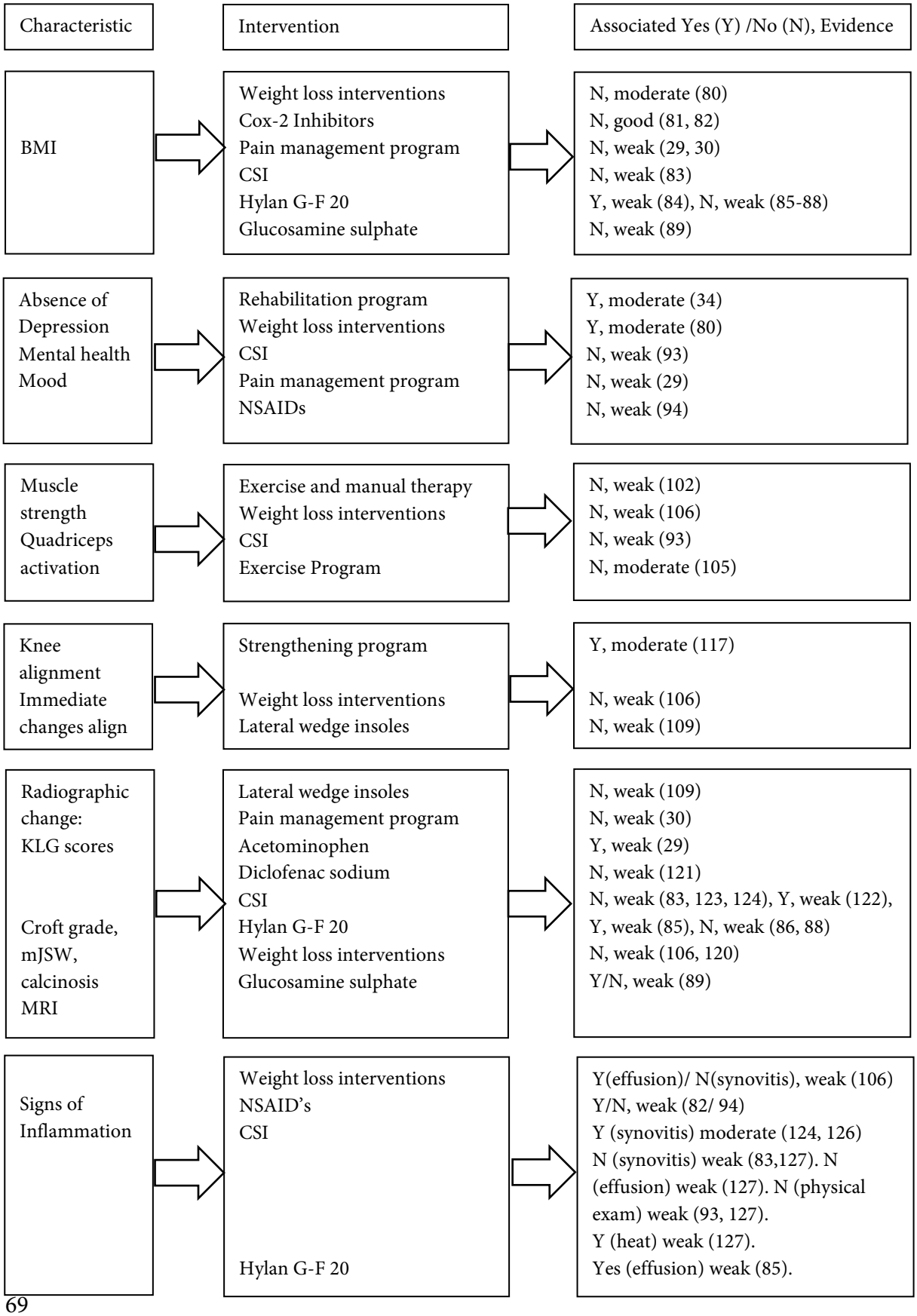
This review identified and summarized the evidence available for particular features of clinical presentation exhibited by people with hip and knee OA that were predictive of response to non-surgical interventions. The studies are summarized in Figure 3.1. Good evidence exists that BMI is not predictive of response to COX-2 inhibitors for hip and knee OA (81), and moderate evidence presented that BMI does not predict weight reduction

following weight loss interventions for overweight people with knee OA (80). There is some evidence to suggest that the absence of depressive symptoms predicts successful outcomes from both weight loss interventions in overweight people with knee OA (80) and a 3-4 week rehabilitation program for participants with hip and knee OA (34). Moderate evidence was cited that quadriceps muscle activation was not predictive of improvements in quadriceps strength attained by participants with knee OA during a strengthening program (105). Patients with medial knee OA who were neutrally aligned were more likely than their more varus aligned counterparts to achieve significant pain relief following a quadriceps strengthening program (117). Evidence was lacking for any radiographic or MRI changes that were significant predictors of response to non-surgical interventions, however, patients with knee OA presenting without inflammatory characteristics on US (synovitis) were more likely to experience prolonged benefit from CSI than inflammatory patients (126).

The practice of analysing patient characteristics as potential predictors of response to interventions is becoming increasingly popular. Researchers attempting to identify predictors of clinical response to non-surgical treatments for hip and knee OA require the use of larger sample sizes, or restriction of the number of variables analysed such that 10-15 responders are studied per possible predictor (128). Identification of further characteristics capable of predicting response to intervention would indeed provide clinicians with additional tools to tailor non-surgical care of patients with hip and knee OA according to their clinical presentation.

Figure 3.1 Clinical characteristics associated with response to nonsurgical management

BMI: Body mass index; CSI: Corticosteroid injection; KLG: Kellgren-Lawrence grade; mJSW: minimum joint space width; MRI: magnetic resonance imaging; NSAIDs: Non-steroidal anti-inflammatory drugs



3.11. Addendum

Since the publication of this review (2013), subsequent studies have examined predictors of response to non-surgical treatments for OA. For the purposes of the research questions of this Thesis, we were most-interested in studies concerned with predictors of outcomes following non-surgical, non-pharmacological interventions such as OA self-management and education, exercise and weight loss. We conducted further literature searches and identified three studies concerned with predictors of response to these interventions published subsequent to the review.

The first study by French et al. (2014) based on secondary analyses from an RCT examined patient baseline characteristics as predictors of OMERACT-OARSI responder criteria following a course of physical therapy for hip OA (129). Although four predictors of response were identified (male sex, lower levels of baseline self-reported physical function, pain, anxiety, and depressive symptoms), the low predictive ability of the model did not provide strong evidence that these variables could adequately predict which patients would be responders (129).

In the second study, Skou et al. (2014) followed a clinical cohort that participated in a combined education and neuromuscular exercise intervention for hip and knee OA. The study aimed to determine if improvements in participant pain (visual analogue scale) and quality of life (EQ-5D) were associated with the following:

i) physical performance and self-efficacy at three months, or

ii) change measured in physical performance and self-efficacy at three months (130). Self-efficacy was defined by the authors as "a person's belief in his or her ability to succeed in a given situation" and was measured using the Arthritis Self-Efficacy scale (pain and other symptoms scale), physical function was measured using the 30-second chair stand test (130). The changes in self-efficacy and the 30-second chair stand test from baseline to three months were associated with one-year improvement in pain. The 30-second chair stand test and self-efficacy at three months were also associated with one-year improvement in pain. Self-efficacy at three months was associated with improvement in quality of life at 12 months.

The third study by Taylor et al. (2018), hypothesised that changes in self-efficacy, perceived pain control, and pain catastrophizing mid-way through a multifaceted intervention for hip and knee OA would predict changes in physical functioning measured on the WOMAC function scale at 12 months (131). The intervention included telephone-delivered weight management, physical activity, and cognitive-behavioural pain management for veterans with OA, in addition to patient-specific treatment recommendations delivered to the primary care providers (PCP) in the study. The control group received usual care from their PCPs. Self-efficacy was measured on the Arthritis Self-Efficacy scale (eight-item scale), pain control was derived from two items of the Coping Strategies Questionnaire and catastrophising was assessed using the Pain Catastrophizing Scale. Changes in self-efficacy and perceived pain control partially mediated improvements in physical functioning (compared to usual care controls) at 12 months however, catastrophising did not (131).

It is interesting to note that two of the studies described above found that changes in self-efficacy were associated with changes in symptoms following interventions that combined several components of non-surgical, non-pharmacological care (130, 131). Although these studies involved different interventions, outcome variables and subscales used to measure self-efficacy, changes in self-efficacy were associated with changes in the outcomes. The authors of these studies recommended that interventions shown to support patient self-efficacy (e.g. education, development of self-management skills) should be considered when delivering treatments to people with OA (130, 131). Considering these relatively recent results, it would be helpful to validate these findings in future research and consider targeting self-efficacy in the design of OA interventions. These studies were not concerned with the association between baseline self-efficacy and clinical outcomes, rather changes in characteristics following periods of treatment were examined. Hence, the relationships between baseline self-efficacy and improvements in pain and function following non-surgical, non-pharmacological interventions have not yet been studied and should be considered in future research.

CHAPTER FOUR

This chapter contains the following published peer-reviewed publication:

Eyles JP, Lucas BR, Patterson JA, Williams MJ, Weeks K, Fransen M, et al. Does clinical presentation predict response to a nonsurgical chronic disease management program for endstage hip and knee osteoarthritis? J Rheumatol. 2014;41(11):2223-31.

Chapter 4

‘Does clinical presentation predict response to a non-surgical chronic disease management program for participants with end-stage hip and knee OA?’

Authors

Jillian P Eyles[1,2,3,4], Barbara R Lucas[1,3], Jillian A Patterson[3], Matthew J Williams[1], Kate Weeks[5], Marlene Fransen[6], David J Hunter[2,3,4]

Affiliations: [1]Physiotherapy Department, Royal North Shore Hospital; [2]Department of Rheumatology, Royal North Shore Hospital; Northern Clinical School; [3]Kolling Institute of Medical Research; [4]Institute of Bone and Joint Research; University of Sydney; [5]Physiotherapy Department, Wollongong Hospital; [6]School of Physiotherapy, Faculty of Health Sciences, University of Sydney.

Corresponding author: David Hunter

Statement from co-authors confirming authorship contribution of the PhD candidate

The co-authors of the paper: "**Does clinical presentation predict response to a non-surgical chronic disease management program for participants with end-stage hip and knee OA**", confirm that Jillian Eyles has made the following contributions:

- Conception and design of the research
- Analysis and interpretation of the findings
- Writing of the manuscript and critical appraisal of the content

In addition to the statements above, in cases where I am not the corresponding author of a published item, permission to include the published material has been granted by the corresponding author.

Jillian Eyles Date: 1st March, 2019

As supervisor for the candidature upon which this Thesis is based, I can confirm that the authorship attribution statements above are correct.

David Hunter Date: 1st March, 2019

Chapter Four: Does clinical presentation predict response to a non-surgical chronic disease management program for participants with end-stage hip and knee OA?

4.1. Abstract:

***Objective*:** To identify baseline characteristics of participants who will respond favourably following 6 months participation in a chronic disease management program for hip and knee osteoarthritis (OA).

***Methods*:** This prospective cohort study assessed 559 participants at baseline and following 6 months of participation in the Osteoarthritis Chronic Care Program. Response was defined as the minimal clinically important difference (MCID) of an 18% and 9-point absolute improvement in the Western Ontario and McMaster Universities Arthritis Index (WOMAC) global score. Multivariate logistic regression modelling was used to identify predictors of response.

***Results*:** Complete data were available for 308 participants, those who withdrew within the study period were imputed as non-responders. Three variables were independently associated with response; signal joint (knee vs hip), gender and high level of comorbidity. Index joint and gender were significant in the multivariate model however, the model was not a sensitive predictor of response.

Conclusions: Strong predictors of response to a chronic disease management program for hip and knee OA were not identified. The significant predictors that were found should be considered in future studies.

4.2. Introduction:

Osteoarthritis (OA) is placed within the world's top 10 most disabling conditions (64). According to the recent global burden of disease estimates, musculoskeletal disorders rank second only to mental and behavioural disorders in overall contribution to years lived with a disability (YLD) (132). A large proportion of YLDs attributed to musculoskeletal disorders results from hip and knee OA, estimated at over 17 million YLDs worldwide (132). Treatments for this disabling, prevalent, and incurable disease focus on symptomatic relief. Numerous international evidence-based guidelines for management of hip and knee OA have become available (67-72). There is consistency in most of the recommendations made by the guidelines (73) and agreement that non-surgical management of hip and knee OA should combine both non- pharmacological and pharmacological treatment modalities (67-72). However, the recommendations are numerous and are not arranged systematically to indicate the order of priority in which treatments should be undertaken or which combinations of modalities should be used. Faced with a plethora of choices, it would be helpful for clinicians to be able to base treatment decisions on the identification of specific clinical presentations that foretell greater likelihood of success following implementation of an individual or combination of treatments. In an era where the delivery of quality care is being promoted coupled with finite resources, the ability to predict outcome/s to intervention would allow clinicians to prioritize those who will redeem the greatest benefit.

There is a growing body of evidence for clinical characteristics that predict response to non-surgical interventions for participants with hip and knee OA (38). Four previous studies have

attempted to identify predictors of response to programs involving combinations of non-surgical interventions (29, 30, 33, 34) however consistent predictors of response were not found. All four treatment protocols involved strategies for self-management of OA including dietary advice, two studies provided weight loss advice if indicated (29, 30) and all were of relatively short duration ranging from 3 to 12 weeks (29, 30, 33, 34). To our knowledge studies reporting outcomes or predictors of response to longer duration self-management programs do not exist. The aim of our research was to determine participant characteristics predictive of favourable outcomes following participation in a longer term non-surgical chronic disease management program for hip and knee OA. We hypothesized that it would be possible to predict participants likely to respond to the program using baseline demographic, psychological, disease related and functional performance variables.

4.3. Materials and Methods:

4.3.1. Study design

This observational clinical cohort study followed consecutive participants of the Osteoarthritis Chronic Care Program (OACCP) from two teaching hospitals in New South Wales (NSW), Australia over a period of six months. The OACCP was developed by the Agency for Clinical Innovation Musculoskeletal Network in response to the growing recognition of the need for a non-surgical care program for people awaiting elective hip or knee total joint arthroplasty (TJA). Participants with symptomatic and radiographic hip and knee OA were recruited for the OACCP at Royal North Shore/ Ryde and Wollongong Hospitals from TJA waiting lists or referral by Rheumatologists, Orthopaedic Surgeons and

General Practitioners. This equates to a doctor diagnosis of OA which provides good face validity (133). People with a diagnosis of knee or hip OA were eligible for the OACCP at initial assessment if they had pain in the affected knee/hip on most days of the past month (47). Participants who had completed a reassessment at 26 weeks (within 140-225 days following initial assessment) were included in the analysis (see Figure 4.1). There were no exclusion criteria for the OACCP, however, participants who did not return for their 26 week assessment, or who were reassessed outside 140-225 days following initial assessment, were considered for imputation as non-responders. Participants imputed as non-responders included; those who underwent TJA more than 90 days (and less than 225 days) following initial assessment, those discharged on medical advice or participants who cited dissatisfaction with the program as their reason for discharge. Those receiving TJA within 90 days of initial assessment were excluded from analysis on the basis that there was insufficient time to determine if they responded to the OACCP. The remaining participants without a complete 26-week assessment within 140-225 days were excluded from the regression analysis.

4.3.2. Intervention

The OACCP aimed to reduce pain, increase function and quality of life of participants with hip and knee OA through provision of access to relevant health professional to support self-management and long term behaviour change. At initial assessment the Musculoskeletal (MSK) Coordinator engaged participants to set goals around the management of their OA and comorbidities (134). The MSK Coordinator was a specialised MSK Physical Therapist; all

participants were prescribed an individualised exercise program that focused on strengthening muscles around affected joints, increasing physical activity levels and other exercises depending on clinical presentation. These programs were reviewed and progressed at 12, and 26 weeks into the program. All participants were provided with education about their OA and any identified comorbidities.

Following initial assessment participants were referred to members of the multi-disciplinary team (MDT) according to clinical need. If participants required medication review they were referred to a rheumatologist or pain clinical nurse consultant. Intra- articular injections were not part of the treatment provided. A dietitian provided interventions when indicated to assist participants with weight loss and/or comorbidity management. Participants requiring assessment of efficiency and safety of functional tasks were referred to an occupational therapist. Psychosocial interventions and linkage with community support services were provided by a social worker as required. Some participants with tibiofemoral or patellofemoral joint malalignment were referred to an orthotist for application of knee bracing or foot orthoses. Participants were also referred to health care providers outside the MDT for other interventions (e.g. hydrotherapy, diabetes education, psychology etc.) as required.

4.3.3. Outcome measures

During a structured interview at initial assessment, the MSK co-ordinator recorded demographic and comorbidity data. Demographic data included: gender, date of birth,

referral source, residential status, language spoken at home, employment and level of education. Signal joint, the predominant site of OA, was determined by clinical examination, patients' symptoms and radiographic evidence of disease. All physical measures performed at initial and 26-week assessments were performed using a standardised protocol (47) including: height, weight, waist and hip circumferences and Body Mass Index (BMI). Disease-specific self-report measures administered at 0 and 26 weeks included the Hip Dysfunction and Osteoarthritis Score (HOOS) or Knee Injury and Osteoarthritis Score (KOOS) according to the signal joint. The Depression Anxiety Stress 21 Scale (DASS 21) was used to measure these 3 negative emotional states at initial and 26 week assessments. The Six Minute Walk Test (6MWT) was completed at baseline and 26 weeks.

HOOS, KOOS and WOMAC

The validated, disease specific HOOS (135) and KOOS (136) require participants to use 5-point Likert scales to rate their: Symptoms, Stiffness, Pain, Physical Function, Recreational Activities and Quality of Life. The HOOS and KOOS subsume all of the WOMAC questions enabling conversion into WOMAC scores (137, 138). The WOMAC subscales for pain, stiffness and function were calculated by summation of the numerical responses provided by the WOMAC questions within the HOOS and KOOS. The WOMAC subscores were combined to calculate a WOMAC Global score= 100- (sum of pain + stiffness + function items) × 100/96. Normalised WOMAC Global Scores were used reflecting the convention that 100 indicated no problems and 0 indicated severe problems (78, 137).

Depression, Anxiety and Stress Scale 21 item version (DASS 21)

Using a 4-point Likert scale, the DASS 21 asks participants to rate how much 21 separate statements applied to them over the past week. The DASS 21 provides subscores to indicate the presence or absence of symptoms of depression, anxiety and stress and has previously been shown to predict the diagnostic presence of depression and anxiety in older adults (139). We were concerned primarily with the depression sub scores; with 0-9 indicating no depressive symptoms, 10-13 mild, 14-20 moderate, 21-27 severe and greater than 28 severe symptoms (140). The DASS depression subscores were categorized into low depressive (0-13) versus high depressive (≥14) groups for the regression analyses.

Modified Self-Administered Comorbidity Questionnaire

This questionnaire asks participants 'has your doctor told you that you have any of the following problems' listing 21 commonly reported conditions plus an 'other' category to indicate comorbidities not included on the list. Response is indicated using 'yes' or 'no'. This questionnaire was adapted from 'The Self-Administered Comorbidity Questionnaire' (141) and is scored by counting 'yes' responses to indicate the number of comorbidities experienced by the participant. The number of comorbidities variable was categorised into low (0-2), high (3-5) and very high (≥6) groups for the analyses.

Six Minute Walk Test (6MWT)

The 6MWT is recommended by the Osteoarthritis Research Society International to assess long distance walking and aerobic capacity for participants with hip and knee OA (142).

Participants were asked to walk as quickly as they could for 6 minutes on a flat 25m track with no corners (143) and the distance walked was recorded in metres. Baseline measurement of oxygen saturation, heart rate and perceived exertion (Borg Scale) were taken prior to and at test completion. Participants with respiratory or cardiac concerns had measures taken at one minute intervals during the test which was discontinued for the following: chest pain or discomfort, mental confusion, lack of coordination, dizziness, intolerable dyspnea, leg cramps, extreme muscle fatigue, persistent oxygen saturation < 85%, other clinically warranted reasons.

Pain Visual Analogue Scale

Participants were asked to rate their average pain on the day of assessment using a visual analogue scale (0 indicated no pain and 10 the most pain imaginable). The pain VAS was categorised into low pain (VAS 0-5) and high pain (VAS 6-10) for the regression analyses.

4.3.4. Statistical analyses

Definition of Responder

Participants were dichotomized according to response or non-response at 26 week assessment with respect to treatment based on the notion of Minimal Clinically Important Difference (MCID) which according to Roman Jaeschke can be defined as “the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient's management”(49). The MCID used was first developed by Angst et al (2002) (32) to

reflect the treatment effect considered to be clinically relevant to a comprehensive rehabilitation intervention for participants with OA of the lower extremities. This MCID required a relative change greater or equal to 18% (100 × (change of score/baseline score) and an absolute change of 9 points improvement of WOMAC global scores at 26 week assessment compared to baseline. Using an MCID comprising of both relative and absolute change standardised the amount of improvement required to achieve response across the spectrum of disease severity. Hence participants with very low global WOMAC scores were not classified as responders for small absolute changes in score compared with those whose baseline scores were higher. Participants who demonstrated improvements in WOMAC global scores at 26 weeks of greater or equal to 18% with an absolute change in score greater than or equal to 9 were categorised as responders (32), those who did not were non-responders. By this definition, participants who presented with baseline WOMAC global scores of > 91 (could not achieve a 9-point difference) or 0 or >84 (could not achieve relative difference of 18%) could not meet the requirements of the MCID and hence were excluded from the analysis. Participants censored at their 26 week follow up due to TJA performed at least 90 days after their initial assessment and within the 26 week assessment window (≤225 days) were imputed into the analysis as non-responders. Participants who withdrew from the OACCP due to dissatisfaction with the program or following medical advice were also imputed as non-responders.

Predictor Variables and Power Calculations

The potential predictor variables were chosen following literature review (38) and discussion amongst this study's authors. The MSK Co-ordinators collecting the data at both study sites were blinded to which variables were to be analysed as predictors of response. Eight baseline predictor variables were identified a priori for consideration in the model: BMI, pain VAS, DASS Depression sub score, signal joint, 6MWT, age, gender and number of comorbidities. The power calculation was set to include at least 10 'responders' per predictor variable (53, 144). Previous studies have reported 34- 47% of participants with hip or knee OA may be expected to satisfy responder criteria following non-surgical multi-modal interventions (30, 34). A sample of 267 was considered sufficient to accommodate 8 predictor variables.

Regression Analyses

Univariate logistic regression analyses examined the association between each of the predictor variables and response, continuous variables were categorised when necessary to meet linearity requirements. All variables were entered into a multivariate binary logistic regression model, the least significant predictor was removed at each step of the modelling until only significant variables remained. To control for confounding, when any variables associated with response in the univariate analyses were removed from the model, the regression coefficients of the remaining variables were checked for a change in 10% or more and if so were retained. Testing for interactions was performed by combining variables of interest. SPSS version 21 was used for all statistical analyses.

Ethics approval was granted by the NSW Population and Health Services Research Ethics Committee AUREI Reference HREC/12/CIPHS/63. Cancer Institute NSW Reference Number 2012/08/413.

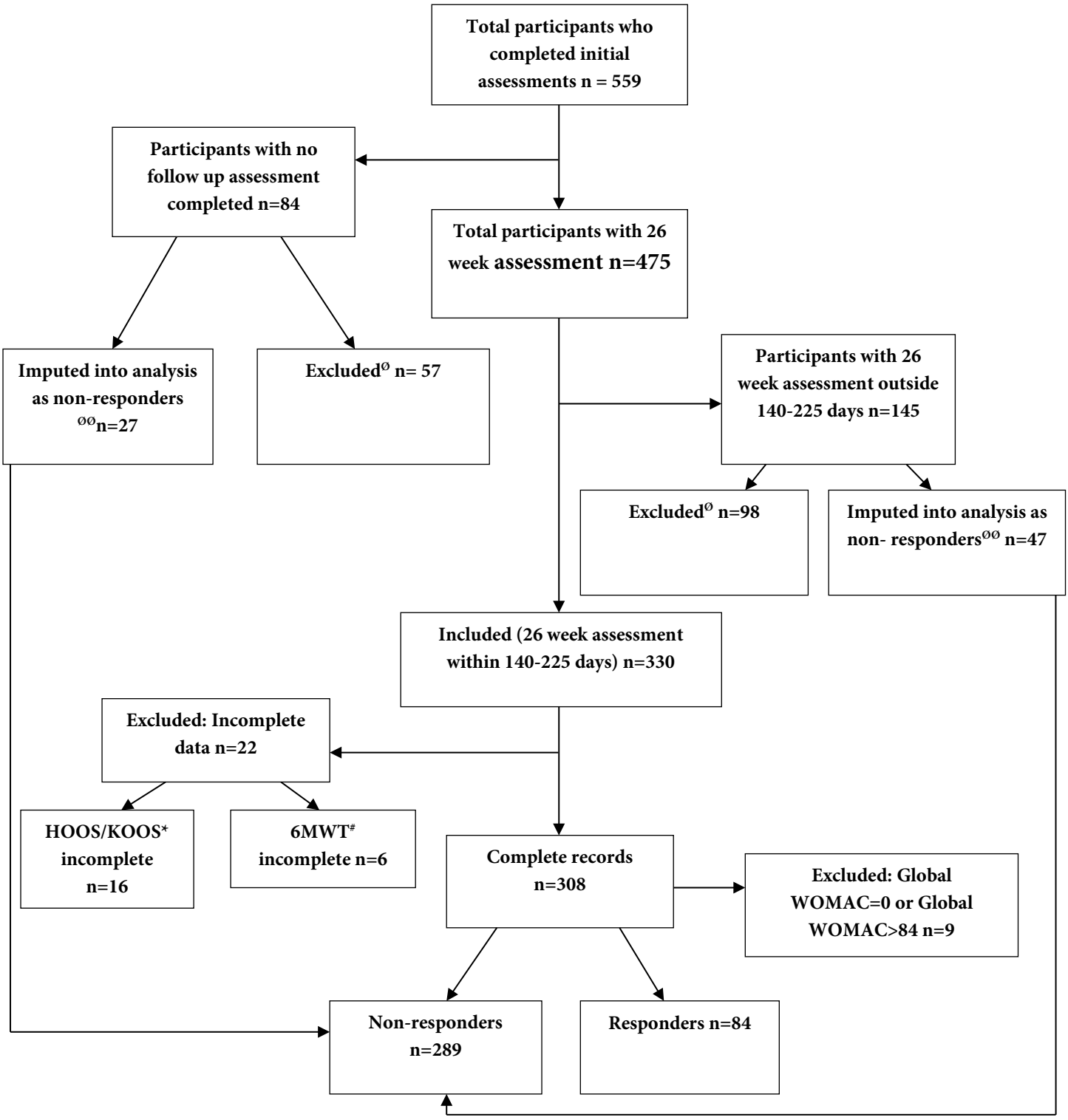
4.4. Results

Of 559 patients consecutively referred to the Wollongong and Royal North Shore/ Ryde Hospitals OACCP July 2011 to December 2013, 475 participants had completed their 26 week assessment as shown in Figure 4.1. There were 145 participants excluded because their 26-week assessment occurred outside the assessment range. A further 16 participants were excluded with incomplete HOOS or KOOS, 6 were unable to complete the 6MWT due to high BP or backpain and 84 did not return for follow up assessment. There were 308 participants with complete datasets remaining for the analysis, a further 74 were imputed as non-responders: 55 discharged from the OACCP after TJA 90-225 days following initial assessment, 16 withdrew due to dissatisfaction with the program and 3 ceased due to medical advice.

The baseline demographics of included participants, those excluded due to missing or assessments outside the 26-week range (n=167) and those who did not return for follow up assessment (n=84) are summarised in Table 4.1. The included and excluded groups were homogenous in most respects. Approximately 90% were referred from elective TJA waiting lists; the wait time for TJA in NSW Hospitals is around 12 months. The majority of participants were of similar age, lived at home with an able person, spoke English, were

retired and overweight. Participants reported similar baseline pain, the majority had 0-5 comorbidities and did not finish high school.

Figure 4.1 Study Flowchart



Ø Participants with incomplete 26 week assessment or 26 week assessment outside 140-225 days or receiving joint replacement surgery within 90 days of initial assessment ØØ Participants who underwent TJA more than 90 days (and less than 225 days) following initial assessment, or were discharged on medical advice or who cited dissatisfaction with the program as the reason for their discharge. *HOOS or KOOS at either 0 or 26 weeks were incomplete so that WOMAC Global scores could not be calculated. #6MWT results were unavailable because participants were unable to complete the test: five due to high blood pressure and one with back pain.

Table 4.1 Participant demographics

Baseline characteristics	Included n = 313	Excluded 26wk not within 140-225 days or missing data n= 162	Excluded no follow-up assessment n= 84	P value [V]
Female (%)	62	59	48	0.067
Age (y), Mean (SD)	68.5 (9.25)	69.0 (9.92)	68.0 (10.85)	0.76
Signal joint knee (%)	77	65	68	0.022
Signal joint knee (%) responders	83			
Signal joint knee (%) non-responders	75			
On Elective Joint Replacement List (%)	88	90	86	0.68
Residence:				
At home with able person (%)	64	68	68	0.46
Home alone (%)	28	22	21	
Other[‡] (%)	8	10	11	
Speaks English[*] (%)	90	92	88	0.61
Employment:				0.60
Not currently employed[†] (%)	86	82	84	
Currently employed[^] (%)	14	18	16	
Education (%)				0.94
Finished secondary school or higher[¤] (%)	30	29	32	
Did not finish secondary school[°] (%)	60	71	68	
BMI (weight(kg)/height (m)^2, Mean (SD)	31.9 (6.88)	32.0 (6.57)	31.7 (6.36)	0.94
BMI knees (weight(kg)/height (m)^2, Mean (SD)	32.52 (7.12)			
BMI hips (weight(kg)/height(m)^2, Mean (SD)	30.03 (5.84)			
Pain VAS, Mean (SD)	5.5 (1.84)	5.7 (1.74)	5.7 (2.20)	0.65
Number of comorbidities				
Low (0- 2) (%)	54	44	43	
High (3- 5)(%)	39	51	42	
Very high (≥6) (%)	8	5	10	
Missing			5	
WOMAC global score[#], Mean (SD), Range	43.4 (19.39), 0- 100	38.4 (17.17), 0- 90	41.3 (21.72), 3- 98	0.027
WOMAC global score knees, Mean (SD)	44.2 (19.66)	40.8 (18.59)	41.5 (21.23)	
WOMAC global score hips, Mean (SD)	40.7 (18.34)	33.6 (12.96)	40.8 (23.23)	
WOMAC global score for 'Responders', Mean (SD), Range	33.8 (18.06), 1-79			
WOMAC global score for 'Responders' knees, Mean (SD), Range	35.6 (18.65), 1, 79			
WOMAC global score for 'Responders' hips, Mean (SD), Range	25.1 (11.71), 6, 47			
WOMAC global score 'Non-responders', Mean (SD), Range	47.4 (18.51), 4- 100			
6 Minute Walk Test (m) Mean (SD)	337.4 (118.52)	324.3 (120.51)	323.5 (114.51)	0.44

[‡] Other includes residence at hostel or residence with non-able person. * Participants who did not speak English (approximately 10%) required the use of an interpreter. [†] Not currently employed includes participants who reported they were retired, performed home duties and other. [^] Currently employed includes participants who reported engaging in full/part-time/ volunteer work. [¤] Includes participants who reported finishing secondary school, tertiary certificate or university graduate. [°] Includes participants who did not finish secondary school, and those who reported no formal schooling. [#] The WOMAC global scores are a transformed score calculated from the HOOS and KOOS, 100 indicates no problems and 0 indicates extreme problems. [V]Independent ANOVA or chi-squared statistic comparing included participants with the 2 other groups.

There were proportionally more males in the excluded group with no follow-up assessment (p=0.07) and the included group reported a higher proportion of knees to hips than the excluded groups (p=0.02). The mean baseline WOMAC global scores were significantly different (p=0.03) however the greatest difference in mean scores was 5.2 points which is not very clinically important.

The referrals to health care providers recorded for included and excluded participants are summarised in Table 4.2. All participants were assessed by a Physical Therapist and provided with a graded exercise program, around half were referred to a dietitian, 30-40% to a rheumatologist and 20-30% to an occupational therapist or a social worker. Approximately 20% of participants were referred to providers within and 40% outside the local health district.

Of 308 included participants with complete datasets, 9 were omitted from analysis because their baseline WOMAC was too high (>84) or zero so were unable to achieve a response. Of the 299 participants with complete datasets, 84 (28%) were responders according to the MCID. Results of the univariate regression analyses are shown in Table 4.3. Compared to females, males were less likely to be responders OR 0.5 (95% CI 0.31, 0.88). There was strong evidence that participants with knee OA were more likely to be responders than those with hip OA OR 2.1 (95% CI 1.10, 3.88). Compared to those with a low number (≤2) there was evidence that participants with a very high number of comorbidities (≥6) were more likely to

Table 4.2 OACCP participant referrals

Healthcare Provider Type	Included participants n=300	Participants without follow up imputed as non-responders n=74	Excluded participants n=185
OACCP Multidisciplinary team			
OACCP Physical Therapist* (%)	100	100	100
OACCP Dietitian* (%)	53.7	55.7	41.3
OACCP Rheumatologist# (%)	40.4	46.3	31.1
OACCP Occupational Therapist* (%)	28.4	36.6	30.5
OACCP Social Worker# (%)	19.4	28.5	13.8
OACCP Orthotist# (%)	23.7	17.8	13.5
Other[t] (%)	16	12.9	10.7
Other health providers within the local health district[±] (e.g. Hydrotherapy, exercise groups) (%)	21	20.3	19.4
Other health providers outside the local health district[V] (e.g. GP, hydrotherapy, diabetes educator, exercise groups) (%)	42	39.2	39.2

* Available at both OACCP sites. # Available at Royal North Shore Hospital OACCP only this rheumatologist saw patients in the OACCP clinic, they did not refer participants to the OACCP. [t] Other may include pain CNC at Wollongong Hospital and education sessions at both sites. [±] Other healthcare providers within the local health district may include hydrotherapy, exercise groups, falls clinic, Physiotherapist, pulmonary rehabilitation, smoking cessation, Geriatrician. [V] Other healthcare providers outside the local health district may include General Practitioner, hydrotherapy, exercise groups, diabetes clinic, Orthopaedic Surgeon, Psychologist, Geriatrician, Physiotherapist, Dietitian, falls clinic, pain clinic, Social Worker, Orthotist, smoking cessation, pulmonary and cardiac rehabilitation.

be responders OR 2.2 (95% CI 0.99, 4.95). The other baseline variables were not independently associated with response.

All potential predictor variables were entered into the base multivariate model. No significant interactions between the variables were found. Following elimination of non-significant variables, the final model (Table 4.4) contained both signal joint ($\chi^2_{LR} = 4.49$, $P < 0.05$) and Gender ($\chi^2_{LR} = 4.95$, $P < 0.05$). Participants with the knee as the signal joint had almost twice the log odds of being responders compared with those with hip OA (adjusted OR 1.92 (95% CI 1.02, 3.62)).

Table 4.3 Univariable analyses

Variable		Unadjusted OR (95% C.I)	P
Age		0.9 (.071, 1.20)	0.539
Gender	Female	Reference	
	Male	0.5 (0.31, 0.88)	0.015
Signal joint	Knee	2.1 (1.10, 3.88)	0.023
	Hip	Reference	
Comorbidity	Low (0-2)	Reference	
	High (3-5)	0.8 (0.47, 1.37)	0.414
	Very high (≥6)	2.2 (0.99, 4.95)	0.053
Depression*	≤13	Reference	
	≥14	1.2 (0.68, 1.98)	0.592
Pain †	0-5	Reference	
	6-10	1.2 (0.72, 1.92)	0.526
BMI‡		1.0 (0.98, 1.05)	0.329
6MWTŁ		1.0 (1.0, 1.0)	0.755

*Depression measured using the Depression component of the Depression Anxiety Stress Scales
†Pain measured using Visual Analogue Scale (self-rated 0 no pain, 10 worst pain). ‡BMI Body Mass Index calculation $\frac{Weight\ (kg)}{Height\ (m)^2}$. ŁDistance participants are able to walk on flat ground during 6-minute walk test

Table 4.4 Multivariable model

Variable	β-Coefficient	P-value	Adjusted OR	95% CI
Constant	-1.496			
Gender	-0.594	0.029	0.55	0.32, 0.94
Signal Joint Knee	0.651	0.045	1.92	1.02, 3.62

The base adjusted or multivariate model included age, gender, index joint, comorbidity, depression, pain, BMI and 6MWT.

Compared to women, men were less likely to be responders adjusted OR 0.55, (95% CI 0.32, 0.94). The very high number of comorbidities group was not significantly associated with response in the multivariate model (p=0.07) and removal did not have a confounding effect on the remaining variables. The model fit the data well using the Hosmer-Lemeshow goodness- of- fit test (χ^2 = 3.03, 3 DF, P = 0.21) however the model was unable to predict any participants as responders (Sensitivity 0%, specificity 100%).

4.5. Discussion

To our knowledge, this was the first study attempting to identify predictors of response following longer term (six months) participation in a chronic disease management program for hip and knee OA. The relatively low response rate (28%) was not surprising considering the severity of disease in this sample indicated by the large proportion of participants on TJA waiting lists (around 90%). Assuming that participants on TJA waiting lists would have clinically and radiographically significant disease, it may be expected that given the natural history of the disease, without intervention the majority of participants would stay the same or worsen over a period of six months. A similar response rate was reported by Weigl et al. (2006) using a less stringent definition of response (≥ 18% improvement in global WOMAC score) six months following a three to four-week rehabilitation program for participants with hip and knee OA (34).

The univariate analysis and the multivariate model adjusting for gender found participants with signal joint the knee had almost twice the log odds of being responders compared to those referred with hip OA OR 1.92 (95% CI 1.02, 3.62). Although signal joint is not a significant predictor of response in the literature (29, 30, 34), this finding makes sense in the clinic. A central aim of the OACCP was to increase physical activity. There is evidence that participants with knee OA experience reduced pain and improvement in physical function following land-based therapeutic exercise (100) however the evidence for such benefits is weaker in those with hip OA (101). Perhaps the participants with knee OA derived higher levels of therapeutic benefit from the exercise prescribed by the Physical Therapist of the OACCP so were more likely to respond than those with hip OA. Included participants with knee OA had a higher mean BMI (32.52 kg/m^2) than those with hip OA (30.03 kg/m^2) seen in Table 4.1. Given that a common goal for OACCP participants was to lose weight, and that participants with knee OA were more overweight, it was hypothesized that these people would be more likely to respond to interventions that involved weight loss. Interestingly BMI was not an independent predictor of response, and it was not significant in the multivariate model when adjusted for the signal joint. This confirms previous findings that BMI was not predictive of responsiveness to weight loss or multimodal non-pharmacological and pharmacological interventions for participants with hip and knee OA (29, 30, 80).

Gender was a univariate predictor of response that remained significant in the multivariate model adjusting for signal joint OR 0.54 (0.31, 0.95). Men had half the log odds of being responders as women, a result that is difficult to explain. The literature yields conflicting

results: female gender was predictive of response to a rehabilitation program for hip and knee OA (34), however gender was not significantly associated with response in other previous predictor studies (29, 30).

Compared to participants with a low number of comorbidities (0-2), participants with a very high number of comorbidities (>6) were independently associated with response OR 2.2 (0.99, 4.95). A very high number of comorbidities was not significantly associated with response when adjusting for gender and signal joint, so number of comorbidities was removed from the model.

The absence of depression has been identified previously as a predictor of response to a 3-4 week inpatient multimodal rehabilitation intervention (34) and positive outcomes from a weight loss program in overweight veterans with knee OA (80). The absence of depression was not a significant predictor of response in the present study. Participants reporting depressive symptoms on the DASS depression subscale were referred for treatment as required. The treatment of depression in people with arthritis has been shown to reduce pain, depressive symptoms, improve function and quality of life (92). The treatment of depression as an adjunct to the other multidisciplinary interventions in our study may have diminished the negative effect depressive symptoms had on response to treatment.

There is a potential limitation regarding the results for DASS depression in this study related to dichotomising the DASS depression variable into low and high depressive symptom

groups. The published clinical thresholds were chosen to characterise the degree of severity of depressive symptoms relative to the population (140). The threshold used to indicate moderate depression or greater was chosen to mark more serious depressive symptoms. Although this method was clinically meaningful, collapsing the groups from five to two categories would have led to 'data loss' which may have limited the results from the analyses of this variable.

Age was not a predictor of response to the OACCP. Most studies include age in their list of potential predictor variables to control for the effects of confounding. Previous evidence for age as a predictor of response is conflicting. Higher age was a predictor of response to a multimodal stepped-care model for participants with hip and knee OA (29) and a Physical Therapy intervention for patients with hip OA (102) however was insignificant in other predictor studies (30, 34). The 6MWT was not predictive of response and while functional performance measures have not been widely used in previous prediction studies, one study found the self-paced 40-metre walk test predictive of response to physical therapy interventions for patients with hip OA (102). A recent systematic review rated the 40-metre walk test as the best walk test based on the limited evidence available (145) and perhaps it would have been a more useful predictor of response for our study. This is an interesting area for future research.

Notable strengths of this study design included: the large sample size, the follow-up period was clinically meaningful and the potential predictor variables were identified a priori through literature and peer review with due consideration to not overfitting the model with

excessive degrees of freedom. The potential predictors included a broad mix of disease, psychological, physical and demographic variables. In order to minimise bias the data was collected prospectively by the MSK co-ordinators who were blinded to which variables were to be analysed as predictors. This clinical cohort study captured data from a real-life clinic. The participants required doctor diagnosis of OA which provides good face validity but may present potential limitations as different symptom labels for OA may exist between independent medical practitioners (133). Recruited largely from TJA waiting lists many participants of the OACCP were censored when their date for TJA came up. Excluded participants reported worse global WOMAC scores at baseline compared to included participants. To control for selection bias, participants who had experienced at least 90 days on the OACCP and had surgery within the 26-week window (≤ 225 days) were imputed as non-responders in addition to those who discontinued the OACCP citing dissatisfaction with the program or withdrew under medical advice. The transformed baseline Global WOMAC score (100 indicates no problems and 0 indicates extreme problems) was significantly lower in responders compared to non-responders ($p<0.05$), and although there is marked overlap between groups, the mean difference of 10 points may suggest some regression towards the mean.

A control group was not used in this study, so it could be argued that it is impossible to distinguish between predictors of response to the chronic disease management program and natural progression of the disease. Previous studies concerned with progression of OA indicate a slow evolution and progression of the disease over time (146). Given that the vast

majority of patients were on the waiting list for TJA indicating end-stage disease, it would be unlikely that the natural course of OA in these participants would allow improvement in symptoms sufficient to achieve the MCID over a period of six months. However this does limit the generalizability of the results of our study to those with severe OA. A previous study reported that compared to participants not waiting for surgery, patients on the waitlist for knee TJA experienced smaller improvements that were not as lasting in response to participation in a chronic disease management program (33). It would be interesting in future research to investigate a more heterogeneous sample of participants to enable analysis of referral for TJA as a potential predictor of response.

We can only assume that referral for TJA was a proxy measure of disease severity in this study. Future research should include a standardised measure of structural disease severity. Higher radiographic severity of knee and hip OA measured using the Kellgren-Lawrence Grading Scale (KLG) was a predictor of response to acetaminophen as part of a Dutch multimodal stepped-care model (29). Conversely, an earlier study investigating predictors of response to the same intervention found that KLG grade was not associated with a more stringent definition of response (30). It would be interesting to investigate whether radiographic severity is associated with response to the longer term chronic disease management program. Another predictor variable in the literature associated with response was a history of previous non-surgical interventions. Two studies reported history of previous non-surgical therapies as associated with good response to rehabilitation programs for participants with hip or knee OA(33, 34), this should be addressed in future studies

concerned with prediction of response to chronic disease management programs for hip and knee OA.

In conclusion, response to intervention could not be predicted using the variables studied in this sample following 6months participation in the OACCP. Although significant predictors of response were identified, the model was not sensitive. The significant predictors of this study should be considered for future research and alternative variables for investigation have been highlighted. It is possible that an alternative battery of variables could be more useful for prediction of response to this intervention. If response can be predicted, it may enable clinicians to better tailor management of hip and knee OA according to clinical presentation.

Acknowledgement

Data for this analysis were obtained with the permission of the Agency for Clinical Innovation (ACI). The authors wish to thank the ACI Musculoskeletal Network especially Mary Fien for assistance with accessing data. We are very grateful to Gary Rolls for his continued support, and Victoria Ireland for assistance with RNSH data collection.

CHAPTER FIVE

This chapter contains the following published peer-reviewed publication:

Eyles JP, Mills K, Lucas BR, Williams MJ, Makovey J, Teoh L, et al. Can We Predict Those With Osteoarthritis Who Will Worsen Following a Chronic Disease Management Program? Arthritis Care Res. 2016; 68 (9): 1268-77.

Chapter 5

'Can we predict those with OA who worsen following a chronic disease management program?'

Authors

Jillian P Eyles[1,2,3], Kathryn Mills[4], Barbara R Lucas[2,3], Matthew J Williams [2,3], Joanna Makovey[1,2], Laurence Teoh[5], David J Hunter[1,2]

Corresponding author: Jillian P Eyles

Affiliations

[1]Department of Rheumatology, Royal North Shore Hospital and Northern Clinical School, University of Sydney, Australia; [2]Kolling Institute of Medical Research, Institute of Bone and Joint Research, University of Sydney, Australia, [3]Physiotherapy Department, Royal North Shore Hospital, Sydney, Australia; [4]Centre for Physical Health, Department of Medicine and Health Sciences, Macquarie University, Sydney, Australia; [5]Department of General Medicine, North Shore Hospital, Takapuna, Auckland, New Zealand.

Statement from co-authors confirming authorship contribution of the PhD candidate

The co-authors of the paper: "**Can we predict those with OA who worsen following a chronic disease management program**", confirm that Jillian Eyles has made the following contributions:

- Conception and design of the research
- Analysis and interpretation of the findings
- Writing of the manuscript and critical appraisal of the content

In addition to the statements above, in cases where I am not the corresponding author of a published item, permission to include the published material has been granted by the corresponding author.

Jillian Eyles Date: 1st March 2019

As supervisor for the candidature upon which this Thesis is based, I can confirm that the authorship attribution statements above are correct.

David Hunter Date: 1st March 2019

Chapter Five: Can we predict those with OA who worsen following a chronic disease management program?

5.1. Abstract

Objective: To identify predictors of worsening symptoms and overall health of the treated hip or knee joint following 26-weeks of a non-surgical chronic disease management program for hip and knee osteoarthritis (OA) and examine the consistency of these predictors across three definitions of worsening.

Methods: This prospective cohort study followed 539 participants of the program for 26-weeks. The three definitions of worsening included: symptomatic worsening based on a change in Western Ontario and McMaster Universities Arthritis Index Global score (WOMACG) measuring pain, stiffness and function, a Transition scale that asked about the overall health of the treated hip or knee joint, and a Composite outcome including both. Multivariable logistic regression models were constructed for the three definitions of worsening.

Results: Complete data were available for 386 participants: mean age 66.3 years, 69% female, 85% knee joint as primary complaint (signal joint), 46% waitlisted for total joint arthroplasty (TJA). TJA waitlist status, signal joint, Six-Minute-Walk-Test (6MWT), depressive symptoms, pain and age were independently associated with at least one definition of

worsening. TJA waitlist status and 6MWT remained in the multivariate models for the Transition and Composite definitions of worsening.

Conclusion: Participants reporting worsening on the transition scale did not consistently meet the WOMACG definition of worsening symptoms. TJA waitlist status was predictive of the Composite definition of worsening, a trend apparent for the Transition definition. However, variables that predict worsening remain largely unknown, further research is required to direct comprehensive and targeted management of patients with hip and knee OA.

5.2. Introduction

Osteoarthritis (OA) is a well-known cause of significant disability (64, 147). Current evidence promotes tailored treatments combining non-pharmacological and pharmacological non-surgical modalities for symptomatic management of knee and hip OA (72, 148, 149). It would be naïve to assume that everybody will benefit from a similar program of non-surgical interventions, hence it is important to identify participants likely to report symptomatic worsening despite 'usual' non-surgical regimens so that alternative therapeutic options may be considered.

Longitudinal studies have examined predictors of hip and knee OA progression: deterioration in radiographic features, symptoms or progression to total joint arthroplasty (TJA) (76, 150-153). Attempts have been made to identify thresholds of pain, function and structural severity to provide a surrogate measure of need for TJA for use in clinical trials (154, 155). Elusive thus far, these thresholds could be useful to triage participants of non-surgical self-management programs so that those who may potentially derive more benefit from TJA may be escalated to surgery. Extant thresholds indicative of symptomatic worsening following a rehabilitation program for hip and knee OA (32) have not been applied extensively in the literature. Yet the ideal threshold for "worsening" may be useful: first, to interpret clinical findings of individual patients and second to derive predictors of worsening in groups of non-surgical program participants. Predictors may be used to triage referrals for interventions unlikely to confer benefit, placing unnecessary burden on the health-care system and patients.

In the absence of a robust clinical definition of "worsening" we have compared two definitions existing in the literature. The first is based on a threshold of change in Western Ontario and McMaster Universities Osteoarthritis Index Global score (WOMACG) derived from participants of a rehabilitation intervention (32). The second is the response to a transition question measuring the overall amount and direction of change the individual has undergone regarding their joint following the intervention (50). This method has been suggested following recent questioning of the utility of pre/post patient reported outcomes (PRO's) to measure efficacy of self-management education programs, possibly due to response shift prompted by change in the participants' perspective following engagement in such programs (50). A third definition of worsening was used in this study: a composite outcome including either or both WOMACG and transition definitions. The composite outcome was chosen to reflect the dual importance of the two clinical outcomes and to increase the power of the analysis by combining two outcomes of common aetiology (52).

The objective of this study was to identify baseline participant characteristics predictive of three different definitions of symptomatic worsening following 26-weeks participation in a non-surgical chronic disease management program for hip and knee OA and examine the consistency of predictors for the definitions. Previous studies have identified age, body mass index (BMI) and pain intensity as predictors of radiographic progression and TJA of knee (76, 150-152) and hip OA (152, 153). As such, these were pragmatically chosen as potential predictors of worsening following participation in the OACCP. TJA waitlist status was also selected based on evidence that some patients report deterioration in health status following

≥6 months wait for TJA (156). Additional predictors included: signal joint, gender, number of comorbidities, a functional performance measure (6MWT) and presence of depression, which have all been previously associated with response to non-surgical rehabilitation programs for hip and knee OA (34, 39, 102). We hypothesized that the same participants with similar demographic, psychological, disease-related and functional performance predictor variables would be identified as 'worse' across three definitions of worsening.

5.3. Patients & methods

5.3.1. Participants and data collection

This study comprised a cohort of consecutive participants with symptomatic and radiographic hip and knee OA recruited for the Osteoarthritis Chronic Care Program (OACCP) at Royal North Shore Hospital (RNSH). Participants were recruited directly from RNSH and Ryde Hospital (New South Wales, Australia) TJA waitlists or referral by rheumatologists, orthopaedic surgeons and general practitioners. People with a diagnosis of knee or hip OA were eligible if they reported pain in the affected knee/hip on most days of the past month (39). Data were included from participants who had completed at least 140 days in the program, there were no exclusion criteria. Ethics approval for analysis of OACCP clinical data was provided by the NSW Population and Health Services Research Ethics Committee AUREI Reference HREC/12/CIPHS/63 Cancer Institute NSW Reference Number 2012/08/413.

The objectives of the OACCP were to reduce pain, increase function and quality of life of participants through provision of tailored interventions delivered by a multidisciplinary team including; a physical therapist, rheumatologist, dietitian, occupational therapist, social worker and orthotist. At initial assessment an experienced musculoskeletal physical therapist (MJW) provided participants with education about their OA and associated comorbidities, set patient-oriented goals, prescribed behavioural modification strategies and an exercise program. The exercise program comprised of strength and cardiovascular training and was progressed at 12, 26 and 52-week reassessments. Participants then attended a multidisciplinary clinic for consultation with a rheumatologist and a selection of other health professionals according to their individual clinical needs.

5.3.2. Outcome Measures

Demographic data were recorded at baseline. Signal joint, the predominant site of OA, was determined by clinical and radiographic examination. Anthropometric measures were performed using a standardised protocol (47) including: height, weight, waist and hip circumferences and Body Mass Index (BMI). Participants rated their average pain on the day of assessment using a 10-cm VAS (0 indicated no pain and 10 the most pain imaginable) (157).

The validated, disease-specific Hip disability and Osteoarthritis Outcome Score (HOOS)(135) and Knee injury and Osteoarthritis Outcome Score (KOOS)(158) require participants to rate their: Symptoms, Stiffness, Pain, Physical Function, Recreational Activities and Quality of Life

on 5-point Likert scales. These questionnaires subsume the WOMAC questions enabling conversion: WOMACG= (sum of pain + stiffness + function items) × 100/96 (78, 137).

The Depression, Anxiety and Stress Scale 21 item version (DASS-21) asks participants to rate how much 21 separate statements applied to them over the past week using a 4-point Likert scale. Subscores indicate the presence/absence of symptoms of depression, anxiety and stress (159). Participants were dichotomised into: those with depression sub-scores 0-9 with no depressive symptoms and those ≥10 with signs of depression (159).

A modified version of the Self-Administered Comorbidity Questionnaire (141) quantified the number of comorbidities experienced by each participant. Participants were asked to respond 'yes' or 'no' to: 'has your doctor told you that you have any of the following problems' followed by a list of 21 commonly reported conditions plus an 'other' category. We categorised the number of comorbidities into low (0-1), moderate (2-3) and high (≥4) groups.

The 6MWT is recommended to assess long distance walking and submaximal aerobic capacity (142). Participants were asked to walk as quickly as possible for 6-minutes on a flat 25m track with no corners (143), the distance walked recorded in metres. Measurement of oxygen saturation, heart rate and perceived exertion were taken prior to and at test completion. Participants with respiratory or cardiac comorbidity had measures at 1-minute intervals the test was discontinued if participants reported concerning symptoms.

At reassessments participants were asked a transition question to rate their signal joint health status compared with prior to starting the program: "Compared with when I started this program, my hip/knee has…" on a 7-point scale: "much improved", "moderately improved" "slightly improved", "not changed", "slightly worse", "moderately worse", or "much worse".

5.3.3. Definitions of worsening

WOMAC Global (WOMACG) definition of worsening: The minimal important difference (MID) is the smallest difference in scores of the variable considered to be beneficial or detrimental by participants (160). MID thresholds for worsening of 9.6 points absolute and 21% relative-change in WOMACG were previously determined for a comprehensive rehabilitation intervention for hip and knee OA (32) .We termed this threshold the "WOMACG definition of worsening". A similar threshold has been used previously to indicate meaningful symptomatic worsening of participants on TJA waitlists (156).

Transition definition of worsening: In addition to pre/post-intervention assessment questionnaires, a transition scale is recommended to assess efficacy of self-management education programs (50). Thus, the second definition of worsening was defined by transition scale participant response "moderately worse" or "much worse". In the absence of evidence for an ideal cut-off for change on the transition question that is meaningful to participants (51), we decided that "slightly worse" was not an adequate threshold for worsening in an attempt to ensure that participants were reporting a change that was important to them.

Composite definition of worsening: The final definition of worsening was based on combined criteria of 9.6 points absolute and 21% relative change in WOMACG scores ***or*** "moderately worse" or "much worse "on the transition scale. This was chosen to reflect the dual importance of self-reported worsening of symptoms and self-reported overall deterioration of the signal joint.

5.3.4. Statistical Analysis

Power calculations were based on evidence that 25% of patients waitlisted for TJA worsened ≥9.6 WOMACG points over 6-months (156). Given that only half of the OACCP participants were on TJA waitlists we extrapolated that around 12.5% of the sample would report worsening. Further, all participants received interventions for their OA, we expected that the majority of participants would report no change or improvements following the intervention and therefore estimated 10% of participants would 'worsen'. A sample of 500 participants was considered sufficient to include 3-5 variables in the final model assuming 10 participants reported worsening per predictor variable (53, 144).

As series of regression analyses were conducted in SPSS (Version 22.0, Armonk NY: IBM Corp, USA). For each model, the dependent variable was "worsening" and was based on dichotomisation of participants into two groups: *worse* and *not-worse* using each definition of worsening. Independent predictor variables were identified *a priori* and the physical therapist collecting data was blinded to which variables were analysed as predictors of worsening.

Missing 26-week assessment data from patients who were not lost to follow-up had their 12-week assessment data carried forward, similar to intention-to-treat analysis. Further, the 6MWT results were standardized to ensure the scale (m) was comparable with the other variables (161).

Univariate logistic regression analyses examined the odds of an OACCP participant worsening when each predictor variable was present. Subsequently, multivariate regression models were built for each definition of worsening. Variables exhibiting odds with $p < 0.2$ that trended in the same direction across a minimum of two definitions were included in the base model for all three definitions of worsening. This method enabled interpretation and comparison of results between the composite definition and the single-outcome definitions (WOMACG and transition definitions)(52). The least significant predictor was removed at each step of the modelling until only significant variables remained. The regression coefficients of the remaining variables were checked on removal of each variable from the model and in the presence of a change of 10% or more the variable was retained. Testing for interactions was performed by combining variables of interest. The validity of each model was assessed using Hosmer-Lemeshow goodness-of-fit and the predictive-ability of the model was assessed by calculation of sensitivity and specificity.

5.4. Results

Of 539 participants consecutively enrolled in the OACCP March 2012 to July 2014, 153 were excluded due to missing data or the presence of a floor/ceiling effect whereby they could not

achieve the WOMACG definition (Figure 5.1). Reasons for missing data included: undergoing TJA, medical advice, moving interstate and illness or death.

Included and excluded participants were of similar age, were overweight and reported low to moderate number (0-3) of comorbidities (Table 5.1). Higher proportions of excluded participants were on TJA waitlists, presented with hips as the signal joint and demonstrated depressive symptoms. Fewer excluded participants finished secondary school ($p=0.018$) and their mean WOMACG scores were higher ($p<0.001$). Excluded patients exhibited worse 6MWT results and higher baseline pain-VAS although the mean difference was within measurement error for both outcomes (162). In the multidisciplinary OACCP clinic, most participants (95%) saw a rheumatologist, 75% were referred to a dietitian, 55% to an occupational therapist, 50% to a social worker and 50% to an orthotist.

Figure 5.1 Study Flowchart

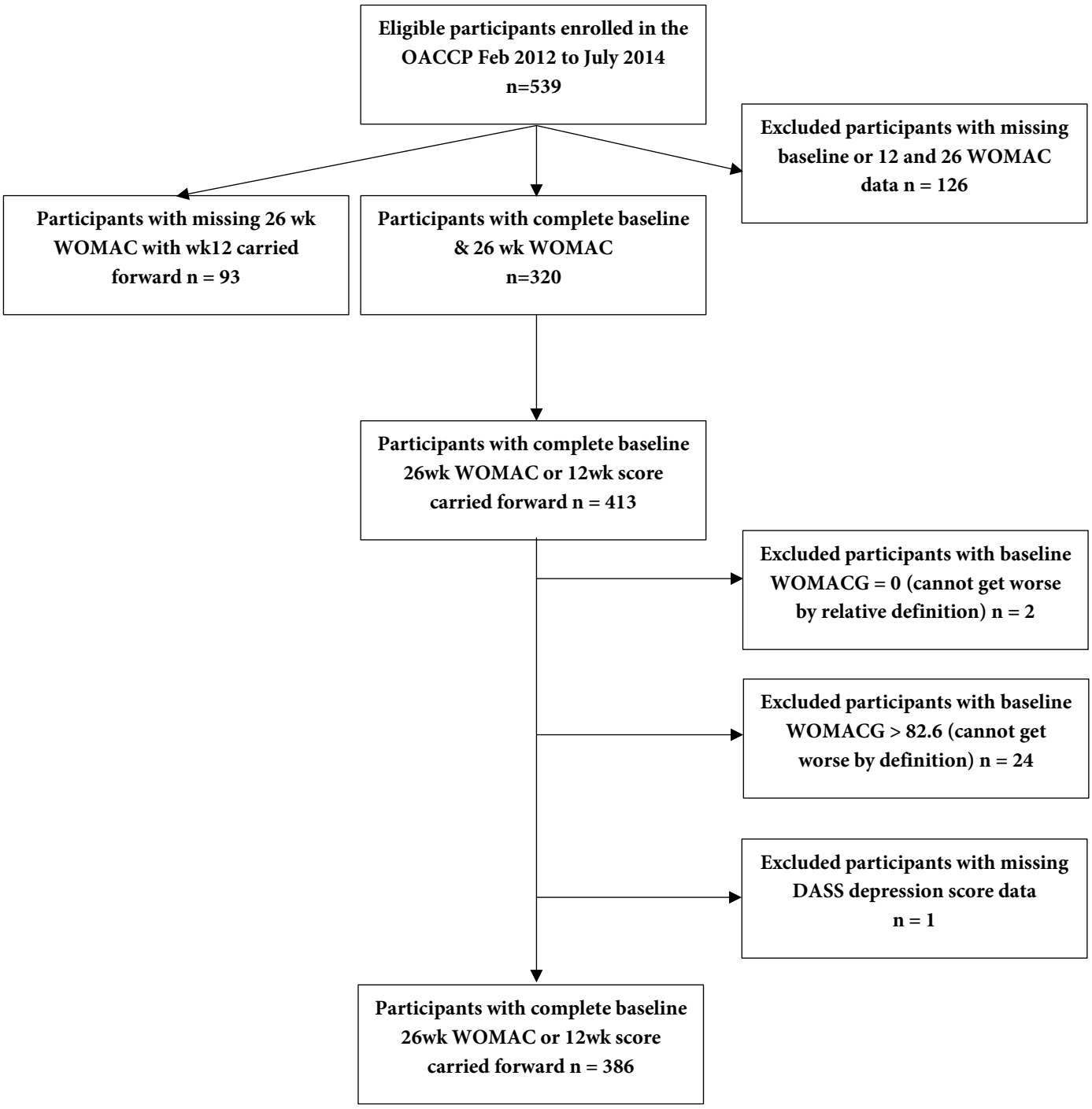


Table 5.1 Participant demographics

Baseline Characteristics	Included n = 386	Excluded: missing data£ n = 153	P[v]
Female (%)	69	63	0.242
Age, years, mean (SD)	66.3 (9.97)	65.8 (11.35)	0.620
Signal joint knee (%)	85	73	**0.006**
On TJA waitlist (%)	46	59	**0.008**
Residence			
Lives alone‡ (%)	28	29	0.694
Speaks English* (%)	93	89	0.146
Engaged in paid employment^ (%)	31	28	0.573
Education			
Finished secondary school or higher∞ (%)	57	46	**0.018**
Did not finish secondary school° (%)	43	54	
BMIº, mean (SD)	30.0 (6.47)	32.1 (7.17)	**0.001**
Pain VAS, mean (SD) range 0-10	4.1 (2.17)	5.4 (2.58)	**<0.001**
6MWT[X], mean (SD)	421.6 (111.72)	380.7 (129.13)	**0.001**
Depressive symptoms: DASS depression subscale (≥14) (%)	35	48	**0.005**
Number of comorbidities			0.233
Low (0-1) %	39	32	
Moderate (2-3) %	39	41	
High (≥4) %	22	27	
Baseline WOMACG#, mean (SD) range	49.1 (18.09) 3.1, 8.3	61.7(23.67) 0, 97.9	**<0.001**
Baseline WOMACG those who got worse vs not worse, mean (SD)			
WOMACG			
Worse, mean (SD) n = 34	38.4 (16.34)		
Not worse, mean (SD) n = 352	50.2 (17.93)		
(CI difference in means), P	(-18.02, -5.45), **<0.01**		
Transition			
Worse, mean (SD) n = 34	58.3 (17.31)		
Not worse, mean (SD) n = 352	48.3 (17.94)		
(CI difference in means), P	(3.69, 16.32), **0.002**		
Composite			
Worse, mean (SD) n = 56	49.3 (19.91)		
Not worse, mean (SD) n = 330	49.1 (17.79)		
(CI difference in means), P	(-4.91, 5.38), 0.929		

Data in **bold face** are statistically significant. £ Missing data include those participants with missing 0, 12 or 26 week WOMACG data (n=126), those whose wk0 WOMACG equalled 0 (n= 2), those with wk0 WOMAC G >82.6 and missing DASS depression data (n=1). ‡ Lives alone reported by participants. Living with others included living with able/non-able bodied person, hostel or aged care residential facility. * Participants who did not speak English (about 7%) required the use of an interpreter. ^ Currently employed includes participants who reported engaging in full/part time paid work. ∞Included participants who reported finishing secondary school (final year), or university degree. ° Includes participants who did not finish secondary school, and those who reported no formal schooling. WOMAC: Western Ontario and McMaster Universities Arthritis Index; # WOMACG: WOMAC Global Scores calculated from the HOOS and KOOS: 0 indicates no problems and 100 indicates extreme problems. WOMACG: WOMAC Global scale minimal clinically important difference for worsening: 9.6 points absolute and 21% relative change in WOMAC Global scores compared to baseline. Transition: transition question definition of worsening; participant response was "moderately worse" or "much worse". Composite: composite definition of worsening; 9.6 points absolute and 21% relative change in WOMAC Global scores compared to baseline **OR** transition question response as "moderately worse" or "much worse". v Independent ANOVA or chi-squared statistic comparing included & excluded participants. VAS: visual analogue scale (cm) ; BMIº: body mass index(kg/m^2). 6MWT[X]: Six minute walk test (m). DASS: Depression Anxiety Stress Scale. TJA: Joint replacement surgery. OR: Odds ratio. CI: Confidence interval.

5.4.1. Definition of worsening outcomes

The 386 participants with complete datasets were included in the regression analyses. Of these 34 (9%) reported worsening according to the WOMACG definition, 34 (9%) met the criteria for worsening for the Transition definition and 56 (15%) for the Composite definition. Only 12 participants met all three definitions of worsening. According to both the Transition OR 2.7 (95%CI 1.26, 5.62) and Composite OR 2.2 (95%CI 1.22, 3.91) definitions, OACCP participants waitlisted for TJA had more than twice the odds of reporting worsening. There was a trend that those with signal joint knees had lower odds of reporting worsening (Table 5.2). The standardised 6MWT was significantly associated with the Transition definition and trended towards significance for the Composite definition of worsening whereby farther walking distance indicated reduced odds of worsening OR 0.7 (95%CI 0.46, 0.99). The results of the univariate analyses differed when the WOMACG worsening definition was applied; participants reporting depressive symptoms were less likely to worsen.

Table 5.2 Univariable analyses for three definitions of worsening

Independent variable	**WOMACG** Unadjusted OR (95%CI)	P	**Transition** Unadjusted OR (95%CI)	P	**Composite** Unadjusted OR (95%CI)	P
Age	1.0 (.99, 1.08)	0.062	1.0 (0.99, 1.10)	0.114	**1.0 (1.00, 1.07)**	**0.028**
Sex						
male	ref		ref		ref	
female	1.1 (0.50, 0.23)	0.895	1.6 (0.78, 3.29)	0.199	1.3 (0.70, 2.28)	0.477
Signal joint						
hip	ref		ref		ref	
knee	1.0 (0.39, 2.84)	0.922	0.5 (0.20, 1.04)	0.063	0.5 (0.27, 1.07)	0.078
Pain VAS	1.0 (0.83, 1.15)	0.820	**1.2 (1.04, 1.45)**	**0.017**	1.1 (1.00, 1.24)	0.203
Number of comorbidities						
low	ref	0.274	ref	0.605	ref	0.365
moderate	1.3 (0.63, 2.88)	0.443	0.9 (0.43, 1.1)	0.845	1.1 (0.60, 2.05)	0.753
high	0.5 (0.17, 1.73)	0.302	0.6 (0.21, 1.68)	0.323	0.6 (0.26, 1.41)	0.244
Depression						
no depression	ref		ref		ref	
any depression	**0.3(0.11, 0.79)**	**0.015**	1.4 (0.66, 2.77)	0.409	0.6 (0.35, 1.22)	0.180
6MWT	1.1 (0.72, 1.5)	0.791	**0.7 (0.46, 0.70)**	**0.033**	0.8 (0.57, 1.03)	0.079
BMI	1.0 (0.95, 1.06)	0.989	1.0 (0.91, 1.02)	0.226	1.0 (0.93, 1.02)	0.351
TJA Waitlist						
on list	ref		ref		ref	
not on list	1.4 (0.67, 2.73)	0.404	**2.7 (1.26, 5.62)**	**0.011**	**2.2 (1.22, 3.91)**	**0.009**

WOMAC: Western Ontario and McMaster Universities Arthritis Index; WOMACG: WOMAC Global scale MCID for worsening: 9.6 points absolute and 21% relative change in WOMAC Global scores compared to baseline. Transition: transition question definition of worsening; participant response was "moderately worse" or "much worse". Composite: composite definition of worsening; 9.6 points absolute and 21% relative change in WOMAC Global scores compared to baseline **OR** transition question response as "moderately worse" or "much worse". ref: reference category for logistic regression. VAS: visual analogue scale DASS: Depression Anxiety Stress Scale. 6MWT: Standardised six-minute walk test. BMI: body mass index (kg/m^2). TJA: Joint replacement surgery. OR: Odds ratio. CI: Confidence interval. P: p-values in bold face if p<0.2, one of the criteria for entry into the multivariate model

5.4.2. Multivariate Models

The final model for the WOMACG definition retained only presence of depressive symptoms OR 0.30 (95%CI 0.11, 0.79) (Table 5.3). The final multivariable model for the Transition definition contained TJA waitlist, 6MWT and signal joint however these predictors did not attain statistical significance (Table 5.3). The final multivariable Composite definition model retained TJA waitlist OR 1.91 (95%CI 1.04, 3.51) and 6MWT OR 0.81 (95%CI 0.60, 1.13).

According to the Composite definition; participants on TJA waitlists had almost twice the odds of worsening. The 6MWT was a confounder; OR 0.8 (95%CI 0.57, 1.03), although as the CI crossed zero, we can interpret this as having little or no effect. The Composite definition model fit the data well using the Hosmer-Lemeshow goodness- of- fit test: (χ^2 = 8.68, 2 DF, p = 0.37), however was only capable of explaining 5% of the variance in demonstrating worsening (Figure 5.2). No model could predict worsening on all 3 definitions together. Although the best model was specific, correctly identifying participants who did not worsen, it lacked sensitivity so failed to identify those who had worsened.

Table 5.3 Multivariable models for three definitions of worsening

Definition of worse	Worse n (%)	Predictors	β-coefficient	P	Adjusted OR	95% CI
WOMACG	34 (9)	Constant	-2.04			
		Depression	-1.21	0.015	0.30	0.11, 0.79
Transition	34 (9)	Constant	-1.14			
		TJA Waitlist	0.65	0.114	1.91	0.86, 4.29
		Signal Joint Knee	-0.54	0.238	0.58	0.24, 1.43
		6MWT	-0. 34	0.092	0.071	0.48, 1.06
Composite	56 (14.5)	Constant	-1.47			
		TJA Waitlist	0.65	0.036	1.91	1.04, 3.51
		6MWT	-0.19	0.231	0.83	0.60, 1.13

WOMACG: WOMAC Global scale MCID for worsening: 9.6 points absolute and 21% relative change in WOMAC Global scores compared to baseline. Transition: transition question definition of worsening; participant response was "moderately worse" or "much worse". Composite: composite definition of worsening; 9.6 points absolute and 21% relative change in WOMAC Global scores compared to baseline **OR** transition question response as "moderately worse" or "much worse". TJA: Total joint arthroplasty6MWT: Standardised six-minute walk test. OR: Odds ratio. CI: Confidence interval.

5.5. Discussion

This study aimed to identify participant characteristics predictive of three definitions of worsening following 26-weeks participation in the OACCP. We hypothesized that the same participants with similar demographic, psychological, disease-related and functional performance predictor variables would be identified as 'worse' across three definitions; however this was not the case. Similar trends were demonstrated by the Transition and Composite models: TJA waitlist and 6MWT were retained in both, though the Transition model also retained signal joint and did not reach statistical significance. These trends were not apparent when the WOMACG definition was applied.

The evidence for symptomatic deterioration while waiting for TJA is conflicting. A systematic review (163) reported people with hip or knee OA waiting for less than six-months did not experience 10% deterioration in WOMAC pain or function scores. In contrast, 25% of people waiting for TJA longer than six-months reported a decline of ≥9.6% WOMACG scores (156). Only 46% of our sample was listed for TJA and these participants had been waiting for approximately six months. This may account for the lack of association between TJA waitlist status and the WOMACG definition of worsening. The participants lost to follow-up due to TJA and other reasons reported a higher mean WOMACG of 61.7 (SD ± 23.67), more similar to the previous study (156). These participants may have worsened in this time without surgery; however we do not have the data to support this supposition. The group reporting worsening according to the WOMACG definition had significantly lower baseline mean WOMACG scores of 38.4 (SD 16.34) compared to 50.2 (SD 17.93) for those who did not

(Table 5.1). One possible explanation is that there was a regression to the mean in participants with lower baseline WOMAC scores.

In contrast to our hypothesis, 22/34 people who reported worsening on the Transition scale were not considered worse according to the WOMACG definition despite using "moderately worse" as the minimum cut-off. This contrasts with the study from which the threshold is derived (32). In the former study participants were not on the TJA waitlist, the intervention focused on physical therapy (not multidisciplinary) and was of shorter duration (3-4 weeks). It is possible that the thresholds we used to indicate worsening were not ideal for our study population. Further work is required to confirm the most appropriate threshold of worsening for this population.

Alternatively, it is possible that discordance of worsening according to WOMACG and Transition scale definitions may be attributed to the attitudes and expectations of participants waitlisted for TJA. The act of booking a person for TJA possibly preconditions them to believe that their signal joint should become worse over time. The language used such as "end-stage", "severe" and "bone-on-bone" may influence their response of "moderately worse" or "much worse" even though their WOMACG scores, a much lengthier questionnaire directly asking about specific symptoms, did not reflect this. The association between use of specific language by care-givers and patient perception of disease severity has received minimal research attention and is a potential area for future work. However, considerable paternalism persists in medical decision making about TJA (164, 165). Perhaps

some of our cohort believed that they needed the surgery because the surgeon said so, and this was sufficient evidence to report that their overall hip or knee joint health was "worse".

Participants on TJA waitlists had twice the odds of meeting the Composite criteria for "worsening" following participation in the OACCP. While TJA can provide good symptomatic relief for most people with end-stage OA (166), non-surgical management is efficacious in reducing the signs and symptoms of knee OA (149). It is advocated that patients be referred for non-surgical treatments as a first port-of-call and participation in chronic disease management programs should commence earlier in the OA disease course prior to being waitlisted for TJA. Although at this stage we can only explain a very small proportion of variables that predict worsening, referring participants earlier in their disease course may lessen their odds of worsening despite taking part in self-management programs.

The absence of depression was a significant predictor of worsening according to the WOMACG definition. This finding is counterintuitive and a possible explanation is that of 34 participants who worsened according to the WOMACG definition, only five reported depressive symptoms. This result is likely to be a type I error. The evidence for the absence of depression as a predictor of positive response to non-surgical interventions for hip and knee OA is conflicting (34, 39, 80). Further investigation of the relationship between symptoms of depression and outcomes following participation in the OACCP is warranted. Depression has been associated with non-compliance with treatment in populations with chronic disease (167). Compliance with OACCP interventions was not measured in this study. Exploration of

the association between compliance with multi-modal therapies, depression and self-reported worsening should also be addressed in future research.

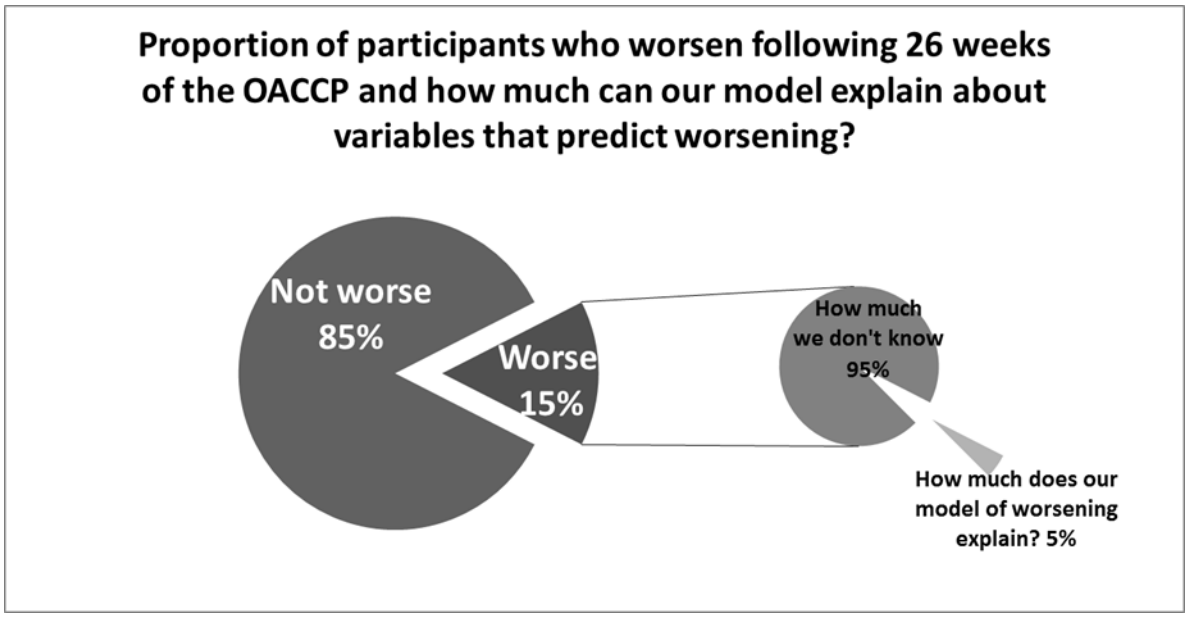
Signal joint was associated with the Transition and Composite definitions of worsening in the univariable analysis; participants with knee OA had lower odds of worsening following the OACCP. However, signal joint was not a significant predictor in the multivariate models. Previous research found that people waiting for hip TJA do not deteriorate compared to those awaiting knee TJA, however those participants were not receiving interventions for their OA (168). In contrast, a previous study following 26-weeks participation in the OACCP those with signal joint knees had twice the odds of responding as those with hips (39). Although exercise for hip OA may confer some reduction in pain and improvement in function, the treatment effect sizes are small (17). Further research into effective non-surgical management options for participants with hip OA is urgently required.

Although previous research reports age, gender and BMI to be important characteristics in disease progression and response to intervention for hip and knee OA (34, 39, 76, 153, 169, 170), they were not found to be significant predictors of worsening in our current investigation. The presence of comorbidities has been associated with poorer health-related quality of life for OA patients (171), yet a greater number of comorbidities was not associated with any definition of worsening. This is an important finding suggesting that non-surgical management may be considered for anyone with any number of concomitant conditions. Increasing pain over time has been associated with progression to TJA (150) and baseline

WOMAC pain scores independently correlated with TJA 6 years following assessment (152). Baseline pain-VAS was only independently associated with the Transition definition, hence failed to meet the criteria for entry into the multivariate models. The 6MWT was not a predictor of "response" in a similar OACCP cohort (38) however the self-paced 40-m walk test was predictive of response to physical therapy interventions for patients with hip OA (102). There is a gap in the research concerning the external validity of the 6MWT as compared to PRO's (145).

This study has several notable strengths: overall it was well-powered, potential predictor variables were identified *a priori* and care was taken to avoid over-fitting the prediction models. A doctor diagnosis of OA was used, which has good face validity (133). Although we were able to determine why participants withdrew participation from the OACCP, WOMAC data was missing for some who progressed to TJA, resulting in their exclusion from the analysis. This potentially limits the applicability of our results. Data from these participants would have been very useful to determine if these participants met any of the 'worsening' definitions immediately prior to their surgery. Significant predictors of worsening were found, however the multivariate models provide a very small proportion of the factors that predict worsening (Figure 5.2) and were not sensitive.

Figure 5.2 Variability of worsening explained by the model



It is possible that stronger predictors of worsening exist that have not yet been studied, or perhaps the thresholds were chosen to represent the predictors affected the outcome. We transformed symptoms of depression and the number of comorbidities into categorical variables which may not have been ideal. This study did not include a control group and was limited to one OACCP site; both investigation into predictors of worsening compared to a control group and use of more heterogeneous samples of participants are important areas for future research.

Three definitions of worsening were applied; potential predictors were identified only when using the Composite definition of worsening. While TJA waitlist status was associated with a two-fold increase in odds of reporting worsening using this definition, the model explained

only 5% of the total variance. Further, following 26-weeks participation in the OACCP, the WOMACG was largely discordant with the Transition and Composite definitions of worsening. Participants with similar demographic, psychological, disease-related and functional performance predictor variables were not consistently identified as 'worse' across the three definitions. Variables that predict worsening are largely unknown and further research into this area is warranted in order to present comprehensive and targeted management of patients with hip and knee OA.

Acknowledgement

Data for this analysis were obtained with the permission of the Agency for Clinical Innovation (ACI). The authors wish to thank the ACI Musculoskeletal Network especially Mary Fien and Robyn Speerin for their roles in the development and oversight of the database and for organising access to the data.

CHAPTER SIX

This chapter contains the following published peer-reviewed publication:

Eyles JP, Hunter DJ, Meneses SRF, Collins NJ, Dobson F, Lucas BR, et al. Instruments assessing attitudes toward or capability regarding self-management of osteoarthritis: a systematic review of measurement properties. Osteoarthritis Cartilage. 2017; 25 (8): 1210-22.

Chapter 6

‘Instruments assessing attitudes toward or capability regarding self-management of osteoarthritis: a systematic review of measurement properties.’

Authors

Eyles, JP[1,2,3], Hunter, DJ[1,2], Meneses, S[1,2], Collins, N[4], Dobson, F[5], Lucas, BR[1,3], Mills, K[6].

Affiliations

[1]Kolling Institute of Medical Research, Institute of Bone and Joint Research, University of Sydney, Australia; [2]Department of Rheumatology, Royal North Shore Hospital and Northern Clinical School, University of Sydney, Australia.;[3]Physiotherapy Department, Royal North Shore Hospital, Sydney, Australia.; [4]School of Health & Rehabilitation Sciences, University of Queensland, Australia.; [5]Centre for Health, Exercise and Sports Medicine, Department of Physiotherapy, School of Health Sciences, The University of Melbourne; [6]Centre for Physical Health, Department of Medicine and Health Sciences, Macquarie University, Sydney, Australia.

Corresponding author: Jillian Eyles

Statement from co-authors confirming authorship contribution of the PhD candidate

The co-authors of the paper: "**Instruments assessing attitudes toward or capability regarding self-management of osteoarthritis: a systematic review of measurement properties.**", confirm that Jillian Eyles has made the following contributions:

- Conception and design of the research
- Analysis and interpretation of the findings
- Writing of the manuscript and critical appraisal of the content

In addition to the statements above, in cases where I am not the corresponding author of a published item, permission to include the published material has been granted by the corresponding author.

Jillian Eyles Date: 1st March 2019

As supervisor for the candidature upon which this Thesis is based, I can confirm that the authorship attribution statements above are correct.

David Hunter Date: 1st March 2019

Chapter Six: Instruments assessing attitudes toward or capability regarding self-management of osteoarthritis: a systematic review of measurement properties

6.1. Abstract

Objective: To make a recommendation on the "best" instrument to assess attitudes toward and/or capabilities regarding self-management of osteoarthritis based on available measurement property evidence.

Methods: Electronic searches were performed in MEDLINE, EMBASE, CINAHL and PsychINFO (inception to 27 December 2016). Two reviewers independently rated measurement properties using the Consensus-based Standards for the selection of Health Measurement Instruments (COSMIN) 4-point scale. Best evidence synthesis was determined by considering COSMIN ratings for measurement property results and the level of evidence available for each measurement property of each instrument.

Results: Eight studies out of 5653 publications met the inclusion criteria, with eight instruments identified for evaluation: Multidimensional Health Locus of Control, Perceived Behavioural Control, Patient Activation Measure, Educational Needs Assessment, Stages of Change Questionnaire in Osteoarthritis, Effective Consumer Scale and Perceived Efficacy in Patient–Physician Interactions five item (PEPPI-5) and ten item scales. Measurement properties assessed for these instruments included internal consistency (k=8), structural validity (k=8), test-retest reliability (k=2), measurement error (k=1), hypothesis testing (k=3)

and cross-cultural validity (k=3). No information was available for content validity, responsiveness or minimal important change/difference. The Dutch PEPPI-5 demonstrated the best measurement property evidence; strong evidence for internal consistency and structural validity but limited evidence for reliability and construct validity.

Conclusions: Although PEPPI-5 was identified as having the best measurement properties, overall there is a poor level of evidence currently available concerning measurement properties of instruments to assess attitudes toward and/or capabilities regarding osteoarthritis self-management. Further well-designed studies investigating measurement properties of existing instruments are required.

6.2. Introduction

Healthcare systems currently face a rising number of people living with chronic conditions leading to disability, without causing death (147). The Chronic Care Model (CCM) has been promoted to assist healthcare systems to meet the escalating demands attributable to chronic conditions (44). The CCM describes healthcare whereby patients are enabled to manage their condition supported by a proactive healthcare delivery system, involving a coordinated team of health professionals with the expertise required to provide decision support, all underpinned by appropriate health information systems (44). Self-management programs are interventions based on the tenets of the CCM; they aim to improve self-management capabilities. It follows that the efficacy of these programs should be measured by assessing change in participants' attitudes toward and/or capabilities to manage their health. However, there are few recommendations guiding which instruments accurately measure self-management (172). The widespread heterogeneity in standardised instruments measuring self-management programs is surprising given that the primary aim of these programs is to directly influence the attitudes toward and abilities to manage one's health.

This situation is apparent in a systematic review that examined the effects of self-management programs for arthritis. Meta-analysis showed short term small to moderate effects for pain up to one year and small long term effects on function for self-management programs for arthritis (14). The studies included in this review examined the effects of self-management programs that typically included the following elements: problem-solving, decision making,

action planning, self-tailoring, pain management strategies, exercise advice, joint protection, rational medication use and physician–patient communication (14).

The studies included in the systematic review of self-management programs focussed on measures of pain and function (14). While these outcomes are obviously important to this population, there appears to be a disparity in the aims of self-management programs and the outcomes used to assess efficacy (15). Self-management programs aim to provide participants with the necessary tools to manage their condition rather than "cure" OA. Although these programs may not dramatically reduce pain and enhance functional ability, this does not necessarily reflect a failed strategy if the participants improve their attitudes towards and ability to manage symptoms and live with an acceptable quality of life despite their disease (15). Another systematic review reported low-to-moderate quality evidence of no or small benefits to participants of OA self-management education programs (15). The authors highlighted the heterogeneity of outcomes used to quantify the effects of self-management programs and that work is needed to establish which outcomes are important to patients. This review recommended rigorous evaluation of OA self-management programs with validated instruments fit to measure attitudes towards/capabilities to self-manage OA, and advised that to achieve this, the measurement properties of the existing instruments need further investigation (15). It is important to note that unlike most of the studies in these systematic reviews, modern OA management programs deliver education and self-management strategies in combination with exercise and weight loss (27, 28, 39). Programs combining these additional treatment modalities should theoretically produce larger

treatment effects, but this is yet to be quantified in a systematic review. Regardless of this, self-management attitudes and capabilities are also important aspects of modern OA management programs and further information regarding the measurement properties of instruments measuring these constructs is required to guide the choice of the 'best' instrument to use.

Measurement properties refer to the ability of the instrument to truthfully and comprehensively measure the specified construct (37). In addition, it is necessary to demonstrate that the instrument is discriminative, sensitive, reliable and deemed feasible in terms of cost and time constraints (36). It is important to consider that the measurement properties of an instrument are not universal across different populations; hence, it cannot be assumed that one with good measurement properties in a specific population will demonstrate the same results in a different population (56). Therefore, the measurement properties of an instrument must be considered within the specific context of the population of interest.

The aims of this systematic review were to: i) identify studies reporting measurement properties of instruments assessing attitudes toward and/or capabilities regarding self-management of OA; ii) systematically critique the studies evaluating instruments using the Consensus-based Standards for the selection of Health Measurement Instruments (COSMIN) tool: and iii) synthesize the evidence available with the possibility of making rudimentary

recommendations concerning the best evidence-based instruments to assess attitudes toward and/or capabilities regarding self-management of OA.

6.3. Methodology

6.3.1. Terminology

Self-management was defined as the individual's ability to manage their physical and psychological symptoms, treatments, consequences and lifestyle changes required to live with their OA (35). Attitudes toward and/or capabilities regarding self-management of OA included the following constructs: knowledge, skills, beliefs, behaviours, activation, self-efficacy, health locus of control, readiness to change healthcare behaviours, healthcare navigation, participation, engagement, and motivation. This list of possible constructs was developed *a priori* using existing content knowledge about available instruments by the authors, and new constructs identified during the review were also included.

6.3.2. Review protocol

The review protocol was developed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement and prospectively registered with PROSPERO on 24 November 2015 (CRD42015019074).

6.3.3. Literature search

The review search strategy was developed and refined by the study authors according to the PRISMA statement and recommendations made for conducting systematic reviews of measurement properties (56, 173). Electronic searches were conducted of the following four bibliographic databases from inception to 27 December 2016: MEDLINE (PubMed), EMBASE (OvidSP), CINAHL (Ebsco), PsychINFO (OvidSP). An initial search was conducted using four main filters containing key search terms as briefly summarised below (see Appendix 1 PubMed search strategy):

I. **Construct**- attitudes toward and capabilities regarding self-management of OA using terms such as: "self-treatment OR self-management OR patient education…" Names of known instruments measuring attitudes and/or capabilities regarding self-management were added using 'OR': "health education impact questionnaire OR patient activation measure OR effective consumer scale …"

II. **Target population**- osteoarthritis OR osteoarth* OR degenerative arthritis OR arthrosis.

III. **Measurement instrument filter**- designed for PubMed to retrieve more than 97% of publications related to measurement properties (59) using terms such as: "instrumentation OR methods OR validation studies…" The filter was translated into the language of the other databases used.

IV. **Exclusion filter**- An exclusion filter was used to improve the precision of the measurement instrument filter (59).

Secondary searching was conducted for all instruments measuring attitudes toward and capabilities regarding self-management of OA identified during the initial search. The name of each instrument was used as the keyword combined (AND) with the target population filter in PubMed. Targeted hand searching of reference lists was also used. Results of the database searches were imported into Endnote X7 (Thomson Reuters, Philadelphia, USA).

6.3.4. Eligibility criteria

Study titles were screened by one reviewer (JE). Two reviewers (JE & SM) independently screened abstracts, followed by the full text of potentially eligible studies. Disagreements were discussed and resolved with a third reviewer (KM). Studies were included if they met the following criteria:

1. **Construct**- at least one instrument attempted to measure the participants' attitudes and/or capabilities regarding self-management of their OA.
2. **Target Population**- adults diagnosed with any stage of OA according to the American College of Rheumatology guidelines, clinical diagnosis of OA from examination findings, patients' symptoms or radiographic evidence of disease. Studies with mixed disease populations were excluded if the proportion of participants with a main diagnosis of OA was less than 80% and the results for OA participants were not reported separately.
3. **Measurement Instrument**- patient-reported outcomes (PROs) (completed by the participant) in the form of questionnaires or scales.

4. **Measurement Properties**- the study was required to explicitly state a primary or secondary aim to develop an instrument or examine at least one measurement property of the instrument involved.
5. **Setting**- the instrument was required to have been utilised in a clinic, field or community setting using readily available equipment. Instruments with a license fee were included.
6. **Publication type**- full-text studies published as original articles in peer-reviewed journals.
7. **Language**- English language studies were included. Non-English language studies were noted, and data extraction performed when possible, however, these were excluded from COSMIN rating due to lack of access to translation resources, and the high level of detail required for a COSMIN review.

6.3.5. Data extraction

Two reviewers (JE & SM) independently extracted data to a predefined spreadsheet with a third reviewer (KM) available to resolve differences. The generalisability of the included studies was considered by extracting characteristics such as mean age, gender distribution, OA stage, setting and language. Relevant data regarding interpretability issues was extracted including the distribution of scores, floor and ceiling effects, change scores, and minimal important change (MIC) or minimal important difference (MID) (174).

6.3.6. Methodological quality evaluation of the studies

Two raters (JE & NC) independently assessed the methodological quality of the included studies, with a third rater (FD) available to resolve discrepancies. Included studies were assessed according to the COSMIN taxonomy of the following measurement properties: internal consistency, reliability, measurement error, content validity, structural validity, hypotheses testing (a form of construct validity), cross-cultural validity, and responsiveness (58). The definitions of these measurement properties are summarised in Table 6.1 (174). Each measurement property featured within a particular study was rated separately according to the COSMIN tool; a robust quality evaluation tool using a 4-point scoring system: "poor", "fair", "good" or "excellent" (60, 174). An overall quality score was given for each measurement property in each study using the "worst score counts" method that accounted for the lowest rating of any item within that measurement property section (60).

Table 6.1 Definitions of measurement properties

Measurement property	Definition
Internal consistency	The degree to which items of an instrument are related to each other
Reliability	The proportion of the total variance of "true differences" measured by the instrument that is not attributed to measurement error
Measurement error	The component of a patient's score that is not due to real changes of the construct measured by the instrument, but attributed to systematic and/or random error
Content validity	The degree to which the content of the instrument measures the construct it intends to measure
Structural validity	The extent to which the scores of an instrument conform to the dimensionality of the construct intended
Hypotheses testing	An aspect of construct validity; when questions are formulated *a priori* about the expected relationships with instruments measuring related constructs
Cross-cultural validity	The extent to which the translated or culturally adapted instrument reflects the performance of the original version of the instrument
Criterion validity	When the scores of an instrument are compared to determine if they are reflective of the outcomes of another instrument considered to be the "gold standard"
Responsiveness	The measurement of the ability of the instrument to detect changes in scores that reflect change in the construct over time
Floor and ceiling effects	The proportion of participants who responded with the lowest or highest possible score on the instrument

Definitions adapted from Mokkink et al. J Clin Epidem 36 (2010) and de Vet, H., et al., "*Measurement in Medicine: A Practical Guide to Biostatistics and Epidemiology" (2010).*

6.3.7. Evaluation of measurement property result

An overall quality rating of the measurement property results for each instrument was performed using a checklist of criteria for good measurement properties (61)(Appendix 2). Two raters determined the quality rating using this additional tool (JE & SM) with disagreements resolved with a third reviewer (NC).

6.3.8. Data synthesis

Qualitative analysis: To summarise the level of evidence of each measurement property for each instrument, a "best evidence synthesis" was performed. The "best evidence synthesis" was derived by triangulating the methodological quality of the studies (174) (using the COSMIN score), the quality criteria for rating the results of measurement properties (Appendix 2) (61), and the level of evidence for the measurement properties of the instruments according to the following: "strong", "moderate", "limited", "conflicting", or "indeterminate" (56, 61); (Table 6.2).

Quantitative analysis: Meta-analysis of data was planned for studies of fair or better methodological quality and of sufficient homogeneity (56).

6.4. Results

The initial search strategy identified 5653 studies (Figure 6.1). Following title and abstract screening, 44 studies were identified for full-text review. Following full-text review, eight

studies were included (175-182). Each study assessed a different instrument, therefore it was not possible to pool data for quantitative analyses. The content of instruments varied widely with respect to the constructs of self-management they represented. Table 6.3 provides a content comparison of the constructs represented in the eight instruments, their characteristics are summarised in Table 6.4. The Patient Activation Measure (PAM) (175) required a license fee; all others were freely available online or following contact with the authors. Many instruments were translated into a language other the original, including Korean (175), Dutch (176, 179-181), Austrian-German, Finnish, Norwegian, Portuguese, Spanish, Swedish (179) and Chinese (182).

Study characteristics such as cohort descriptors, sample sizes and instrument scores are provided in Table 6.4. The OA sites captured within the studies included hand, hip and knee (176, 179), hip and knee (177), knee (182)or were not specified (175, 178, 180, 181). Stage or duration of OA was generally unreported. Participants were predominantly female across all studies and representative of the age of the wider OA population, with mean age ranging from 62-72.2 years.

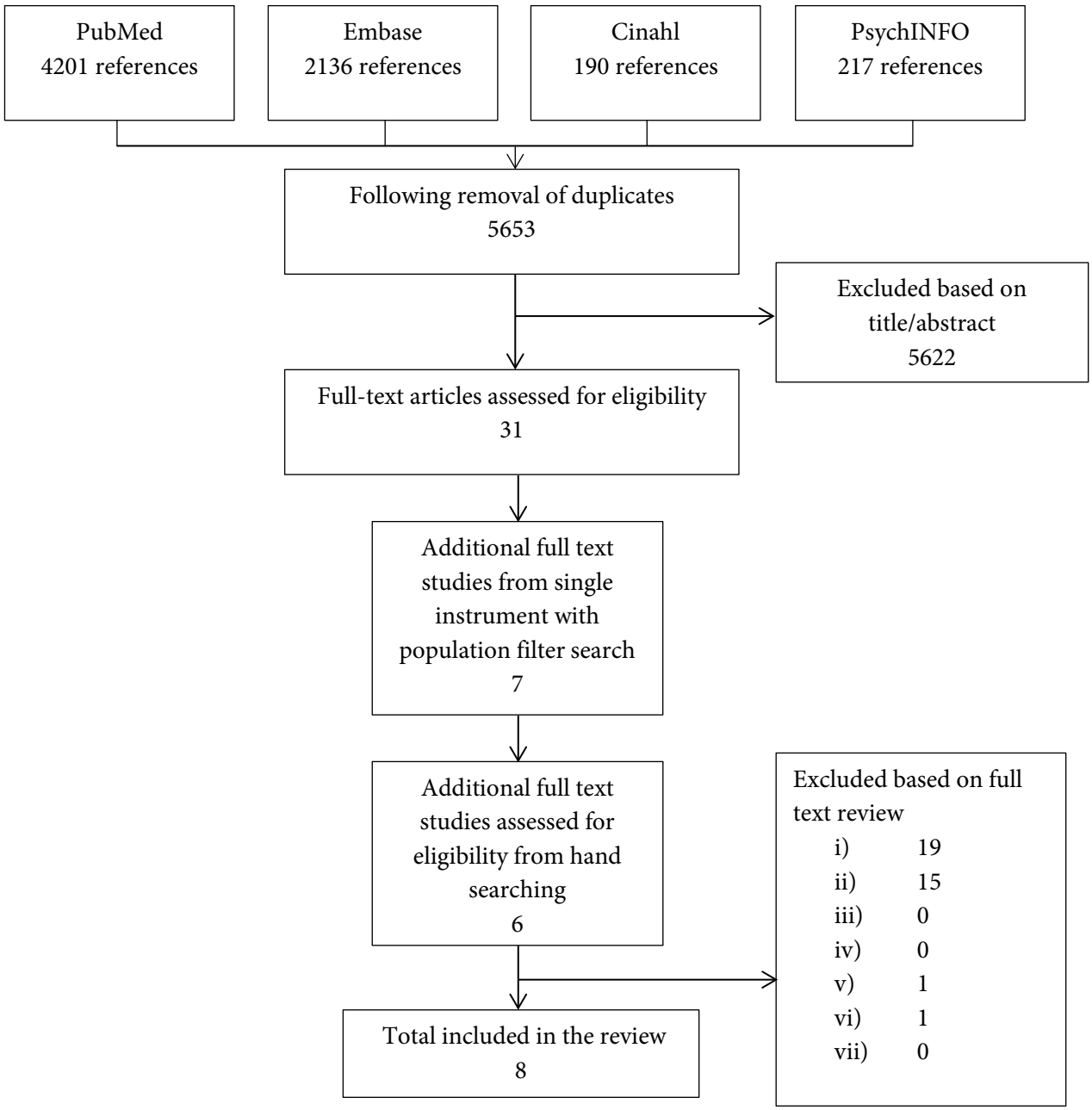
Table 6.2 Levels of evidence for the quality of the measurement property

Level of evidence	Rating	Criteria
Strong	+++ OR ---	Consistent findings in multiple studies of good methodological quality OR in one study of excellent methodological quality
Moderate	++ OR --	Consistent findings in multiple studies of fair methodological quality OR in one study of good methodological quality
Limited	+ OR -	One study of fair methodological quality
Conflicting	±	Conflicting findings
Indeterminate	?	Only studies of poor methodological quality

Adapted from Terwee et al. J Clin Epidemiol 2007;60(1):34-42

+ = positive rating, ? = unknown rating, - = negative rating.

Figure 6.1 Flowchart of the selection and inclusion of studies



Exclusion Criteria
i) Population: the proportion of participants with a main diagnosis of OA was less than 80% and the results for OA participants were not reported separately
ii) Construct: Not an instrument that measures attitudes or abilities pertaining to self-management of OA
iii) Instrument: Not a patient-reported outcome in the form of questionnaire or scale
iv) Setting: Not used in a clinic setting/field
v) Measurement study: No primary or secondary aim to examine at least one measurement property
vi) Publication type: Not a full-text article
vii) Language: Not English (only excluded from COSMIN review)

Table 6.3 Content comparison of instruments

Construct	Attitudes/ beliefs pertaining to self-management of OA	Attitudes/ beliefs pertaining to changing health behaviour	Knowledge required for self-management	Capability to perform skills required for self-management	Educational needs for self-management of OA	Interactions with health care providers assisting with management of OA	Overall capability to self-manage OA
MHLC[173]	•						
PBC[174]	•			•		•	
PAM-13[171]	•		•	•		•	•
ENAT[175]	•	•	•		•		
PEPPI-5[176]			•	•		•	
SCQOA[172]	•	•					
EC-17[177]	•		•	•		•	
PEPPI-10[178]			•	•		•	

MHLC= Multidimensional Health Locus of Control, IHLC= Internal Health Locus of control, PBC= Perceived behavioural control, PAM13= Patient activation measure, ENAT= Educational needs assessment, PEPPI-5= Perceived Efficacy in Patient–Physician Interactions Scale, SCQOA= The Stages of Change Questionnaire in Osteoarthritis, EC-17= Effective Consumer Scale.

Table 6.4 Characteristics of included studies

Authors/ Instrument	Construct described	Time to administer	Availability	Language & country	Number, type of questions & scoring	Proport -ion with OA (%)	OA site & stage	% other diseases in sample	N with > 80% OA (response rate %)	Age: mean age years (SD) or age groups (%)	Female %	Mean (standard deviation), possible score range, distribution
Kelly (2007)/ MHLC[173]	Measures beliefs about who or what controls the patient's health status	Not stated	Freely available at: http://www.nursing.vanderbilt.edu/faculty/kwallston/mhlcscales.htm	English, USA & Canada	Three scales of 6 items each, using 6-point Likert scale measuring the following dimensions: "Internal" "Chance" and "Powerful Others". Sum the individual item scores for each subscale.	86.2	Hip & knee	Control sample: 13.8	1040 (100)	Study I: 65 (9) Study II: 64 (16) Study III: 62 (6)	Study I: (66) Study II: (59) Study III: (63)	IHLC: 26.44 (5.61) PHLC: 20.22 (6.64) CHLC: 16.96 (6.05) Each subscale has range 6- 36
Liu (2007)/ PBC [174]	Survey of OA patients' drug information seeking from physicians and pharmacists.	Not stated	In published paper	English USA	8 statements with 7-point Likert responses Perceived difficulty:3 Self-efficacy: 3 Controllability: 2 Answer for physicians & pharmacists separately	100	Not stated	-	1000 (61.9)	18-24: 1.8% 25-34: 3.8% 35-44: 11.9% 45-54: 27.6% 55-64: 28.3% >64: 26.6%	72.8	PDP: 5.10 (1.60) PDPh: 5.27 (1.49) SEP: 5.62 (1.62) SEPh: 5.62 (1.60) CP: 5.63 (1.36) CPh: 5.62 (1.37)
Ahn (2015)/ PAM-13[171]	Patient activation: patient's knowledge, skill, and confidence regarding the self-management of a chronic disease	Not stated	Insignia Health provides licenses for the PAM at a cost	Korean, South Korea	13-statements, with responses on a 4-point Likert scale. Raw score: sum responses to the 13 items. Scores ranging from 13 to 52. converted to a 0–100 interval scale. Higher total PAM scores reflect higher levels of patient activation.	100	Not stated	-	270 (100)	72.2 (8.3)	82.4	50.0 (13.5) 0-100,

Ndosi (2014)/ ENAT[175]	Assesses the educational needs (priorities) of patients with rheumatic diseases	Not stated	Contact authors	Austrian German Finnish Dutch Norwei-gian Portu-guese Spanish Swedish Austria Finland Nether-lands Norway Portugal Spain Sweden	39 items with 4-point Likert scale in 7 domains: managing pain (6 items), movement (5 items), feelings (4 items), arthritis process (7 items), treatments (7 items), self-help measures (6 items) and support systems (4 items)	14.4	Hand, hip or knee in discussi on. Stage not stated	AS: 22.5% FM: 12% PsA: 26.8% SLE: 12.3% SS: 12.0%	433 (response rate not stated)	Not stated for OA sample: pooled sample is 52.6(13.1)	Not stated for OA but across pooled sample 66.2	Not stated for OA group
ten Klooster (2012)/ PEPPI-5[176]	Self-efficacy in both obtaining medical information and attention to chief health concern from a physician	Not stated	Dutch version freely available on web. English version published	Dutch, Nether-lands	5 questions with responses on a 5-point numerical rating scale. Total scores are summed to range from 5 to 25, higher total scores reflect higher perceived self-efficacy in patient–physician interactions.	100	Not stated	-	224 (55.4)	62.9 (10.2)	81.3	18.8 (4.3) 5- 25, Slightly negatively skewed.
Heuts (2005)/ SCQOA[172]	People move from low to high level of participation. Stages: no intention to change to optimal	3-5 min	Published in paper as appendix (in English)	Unclear (Dutch or English), Nether-lands	21 items scored on 5-point likert scale. 3 subscales: 7 questions for precontemplation, 7 for contemplation, 7 for action.	100	In results hip, knee & hand. Stage not stated	-	273 (100)	Range 40-60 years for inclusion criteria	59.7	Using highest score method: 10.3% was in the 'pre-contemplation stage', 22.3% in the

	active participation with internalization of new behavior											'contemplation stage', 67.0% was 'in action',
ten Klooster (2013)/ EC-17[177]	Measures knowledge, attitudes, and behaviours regarding self-management skills	Not stated	Available in published paper & on web http://www.cgh.uottawa.ca/assets/documents/Survey.pdf	Dutch Nether-lands	17 items with 5-point Likert scale. Item scores are summed when items are completed and converted to range from 0 to 100, where 100 is the best possible score.	85.6	Not stated	FM: 14.4	209 (55.8% of combined OA & FM sample)	62.6 (10.1)	80.9	68.9 (16.3), 0-100, Near normal distribution (Kolmogorov-Smirnov, P= 0.058)
Zhao (2016) PEPPI-10[178]	Self-efficacy in both obtaining medical information and attention to chief health concern from a physician	Not stated	Supplement link from paper: https://www.dovepress.com/get_supplementary_file.php?f=110883.pdf	Chinese, China	10 items with 10 point Numerical Rating Scale: Not confident to extremely confident. Sum ten scores from 0 to 100 (100 best self-efficacy)	100	Knee	-	115 (100)	63.42 (6.7)	59	90.07 (12.9), 0- 100 Negatively skewed distribution

MHLC= Multidimensional Health Locus of Control, IHLC= Internal Health Locus of control, PHLC= Powerful Others Health Locus of control, CHLC= Chance Health Locus of Control, PBC= Perceived behavioural control, PDP= Perceived difficulty for physicians, PDPh= Perceived difficulty for pharmacists, SEP= Self-efficacy for physicians, SEPh= Self-efficacy for pharmacists, CP= Controllability for physicians, CPh= Controllability for pharmacists, PAM13= Patient activation measure, ENAT= Educational needs assessment, PEPPI-5= Perceived Efficacy in Patient–Physician Interactions scale, SCQOA= The Stages of Change Questionnaire in Osteoarthritis, EC-17= Effective Consumer Scale. RA: Rheumatoid arthritis, FM: Fibromyalgia, AS: Ankylosing spondylitis, PsA: Psoriatic arthritis, SLE: Systemic Lupus Erythematosus, SS: Systemic sclerosis.

6.4.1. Measurement property results and "best evidence synthesis"

Findings for measurement properties are summarised in Tables 6.5 and 6.6, qualitative data synthesis in Table 6.7.

Internal Consistency

Internal consistency was estimated for all instruments. Strong evidence (excellent rating) for internal consistency (Cronbach's α = 0.92) was found for the Perceived Efficacy in Patient–Physician Interactions 5 item scale (PEPPI-5) (180), satisfying requirements for unidimensionality (Appendix 2). Moderate evidence (good rating) of adequate internal consistency was demonstrated for the Perceived Efficacy in Patient–Physician Interactions 10 item scale (PEPPI-10) (182) (Cronbach's α = 0.91). Limited evidence (fair rating) of adequate internal consistency was found for three instruments: Perceived Behavioural Control (PBC) (178), PAM-13 (175) and The Stages of Change Questionnaire in Osteoarthritis (SCQOA) (176). There was indeterminate evidence (poor rating) of internal consistency for three instruments: Multidimensional Health Locus of Control (MHLC) (form C) (177), Educational Needs Assessment Tool (ENAT) (179) and Effective Consumer Scale (EC-17) (181).

Table 6.5 Measurement property results: internal consistency, reliability, measurement error and structural validity.

Instrument	#Requirements IRT	Internal consistency		Reliability		Measurement error		Structural validity	
		Result Cronbach's alpha	COSMIN score	Result	COSMIN score	Result	COSMIN score	Result	COSMIN score
MHLC[173]	Good	IHLC: 0.75; PHLC: 0.70; CHLC: 0.65	Poor	-	-	-	-	Confirmatory FA, 3 factor model: χ^2= 904.50 , 135 df, (P<0.01), RMSEA 0.0, GFI = 0.96, CFI= 0.79, ECVI= 0.81,PCA, FA & Rasch analysis supported item reduction: removed 2 items	Poor
PBC[174]	-	PDP: α= 0.77 PDPh: α= 0.72 SEP: α= 0.83 SEPh: α= 0.83	Fair	-	-	-	-	PCA & exploratory FA with Factor loading. Data reduction & data detection	Fair
PAM-13[171]	Good	α= 0.88	Fair	-	-	-	-	Confirmatory PCA GFI= 32 (11.9%) misfits MNSQ 0.68 to 1.42 Rasch analysis: person reliability was between .87 (real) and .89 (model), and the item reliability was .99. The separation index for persons was 2.57 and that for items was 10.56. 57.5% variance of data explained.	Fair

ENAT[175]	Good	IRT: Person separation index > 0.9	Poor	-	-	-	-	Confirmatory FA, structure detection & Rasch analysis OA group was a misfit	Fair
PEPPI-5[176]	-	α= 0.92	Excellent	Test-retest: ICC 0.68 (95% CI 0.56, 0.78) Bland–Altman analysis LOA 6.83 -6.35 (mean difference -0.24, t(99) = -0.71, p = 0.48)	Fair	LOA -6.83-6.35 differences _ weakly related to the magnitude of the measurement (r2 = 0.04, p = 0.049), indicating little to no systematic bias.	Fair	Confirmatory FA, factor loading & structure detection (1 factor) SB χ^2 (5) = 17.43, NNFI = 0.98, CFI = 0.99, SRMR = 0.03, RMSEA (90% CI) = 0.11 (0.05–0.16)	Excellent
SCQOA[172]	-	action α= 0.74 precontemplation α= 0.70 contemplator α= 0.77 After removal of 5 items: action α= 0.79 precontemplation α= 0.72 contemplation α= 0.76	Fair	-	-	-	-	Confirmatory FA, factor loading & date reduction: removal of items 3, 7, 12, 16, 18 and 20 PCA. Repeated FA with 15 item scale: 3 factors explained 45% of variance	Fair

EC-17[177]	Good	person reliability: 0.92	Poor	test-retest ICC= 0.71 (95 % CI: 0.60–0.80)	Poor	-	-	Confirmatory FA Apart from RMSEA, 1-factor model good fit SB χ^2 (119) = 488.70, NNFI = 0.96, CFI = 0.96, SRMR = 0.08, RMSEA (90 % CI) = 0.11 (0.10–0.12).	Poor
PEPPI-10[178]	-	α= 0.91	Good	-	-	-	-	Confirmatory FA: two-factor model good fit (df=33, P-value =0.000) except RMSEA=0.164 above cutoff	Good

NOTE: Content validity, criterion validity and responsiveness were not reported on in any included articles, hence do not appear in the table.
[#]This field was only completed for those instruments based on Item Response Theory (IRT). MHLC= Multidimensional Health Locus of Control, IHLC= Internal Health Locus of control, PHLC= Powerful Others Health Locus of control, CHLC= Chance Health Locus of Control, PBC= Perceived behavioural control, PDP= Perceived difficulty for physicians, PDPh= Perceived difficulty for pharmacists, SEP= Self-efficacy for physicians, SEPh= Self-efficacy for pharmacists, CP= Controllability for physicians, CPh= Controllability for pharmacists, PAM13= Patient activation measure, ENAT= Educational needs assessment, PEPPI-5= Perceived Efficacy in Patient–Physician Interactions 5 item scale, SCQOA= The Stages of Change Questionnaire in Osteoarthritis, EC-17= Effective Consumer Scale. FA= Factor Analysis, PCA= Principal Components Analysis, GFI= Goodness of fit index, MNSQ= Infit & outfit mean square statistics, NRS= numerical rating score, NS= non-significant, NNFI= Non-normed Fit Index, CFI= comparative fit index, SRMR= standardized root mean square residual, RMSEA= root mean square error of approximation, SB χ^2= Satorra-Bentler chi-squared statistic, LOA = limits of agreement, MFES= modified fall efficacy scale, OSES= osteoporosis self-efficacy scale, PEPPI-10= Perceived Efficacy in Patient–Physician Interactions 10 item scale; SEE-C= self-efficacy for exercise scale.

Table 6.6 Measurement property results: construct validity, cross-cultural validity, and floor and ceiling effects.

Instrument	Construct validity (Hypothesis testing)			Cross-cultural validity		Floor & ceiling effects
	Hypothesis	Result	COSMIN score	Result	COSMIN score	Result
MHLC[173]	-	-	-	-	-	Seven items, including all six items of the IHLC scale, exhibited skewness that exceeded -1.00 (i.e., a "ceiling effect"). No floor effect.
PBC[174]	-	-	-	-	-	-
PAM-13[171]	-	-	-	Items 1 and 4 were adjusted to make more sense in Korean translation. PCA indicated unidimensionality	Fair*	-
ENAT[175]	-	-	-	-	-	-
PEPPI-5[176]	Expected correlations: Strongly positively correlated with EC-17, moderately positively with GSES, weakly positively with AIMS2 family & friends scale and SF-36 MCS and not correlated with SF- 36 PCS and pain NRS	EC-17: r=0.52, p<0.01 GSES: r= 0.07 (not sig) AIMS2 F & F: r=0.23, p<0.05 SF-36 MCS: R= 0.26, p<0.01 SF- 36 PCS: r= 0.05 (NS) Pain NRS: r=-0.12 (NS)	Fair	-	-	No floor and ceiling effects: no patients scored five and 26 patients (11.6%) scored 25.

SCQOA[172]	-	-	-	-	-	-
EC-17[177]	Expected correlations: Strongly correlated PEPPI-5, moderately correlated GSES and AIMS2 F & f, moderate correlation SF-36 MCS, weak correlations SF-36 PCS & pain NRS	PEPPI-5: r=0.55, P<0.01, GSES: r=0.26, P<0.01 AIMS2 F & F: r=-0.34, P<0.01 SF-36 MCS: r=0.39, p<0.01 SF-36 PCS: r=0.14, p<0.05, Pain NRS: r=-0.21, p<0.01	Poor	Following pretests small wording changes made in six items. CFA supported unidimensional structure of the scale	Poor*	No ceiling or floor effect found: no participants scored zero and only 1.3% achieved maximum score
PEPPI-10[178]	No hypothesis and expected correlations not stated	SEE-C: r=0.292, p<0.01, MFES: r= 0.220, p<0.05, OSES: r=0.315, p<0.01	Poor	Following pretests, two items were modified to suit Chinese language. FA showed Chinese version of PEPPI-10 has two common factors; different to 1 factor reported previously for the English version.	Fair*	Ceiling effect found for 28.2% of participants. No floor effect.

NOTE: Content validity, criterion validity and responsiveness were not reported on in any included articles, hence do not appear in the table. Floor and ceiling effects were not evaluated using the COSMIN Checklist.*Paper did not assess cross-cultural validity however did translate the questionnaire into other language(s) hence quality of translation items of COSMIN checklist were rated (Box G items 4-11).MHLC= Multidimensional Health Locus of Control, IHLC= Internal Health Locus of control, PBC= Perceived behavioural control, PAM13= Patient activation measure, ENAT= Educational needs assessment, PEPPI-5= Perceived Efficacy in Patient–Physician Interactions 5 item scale, SCQOA= The Stages of Change Questionnaire in Osteoarthritis, EC-17= Effective Consumer Scale, GSES= General Self Efficacy scale, AIMS2 F & F= Dutch Arthritis Impact Measurement Scales 2 Family and Friends scale, SF-36 MCS= Short form 36 mental component summary score, SF- 36 PCS= Short form 36 mental component summary score, MFES= modified fall efficacy scale, OSES= osteoporosis self-efficacy scale, PEPPI-10= Perceived Efficacy in Patient–Physician Interactions 10 item scale; SEE-C= self-efficacy for exercise scale.

Table 6.7 Measurement property synthesis using COSMIN rating, quality criteria and levels of evidence

Instrument	Internal consistency	Reliability	Measurement error	Structural validity	Hypothesis testing	Cross-cultural validity	Floor and ceiling effects
MHLC[173]	?	0	0	?	0	0	?
PBC[174]	+	0	0	?	0	0	0
PAM-13[171]	+	0	0	+	0	*+	0
ENAT[175]	?	0	0	-	0	0	0
PEPPI-5[176]	+++	-	?	+++	+	0	+++
SCQOA[172]	+	0	0	-	0	0	0
EC-17[177]	?	?	0	?	?	*?	?
PEPPI-10[178]	++	0	0	++	?	*+	-

NOTE: Content validity and responsiveness were not reported on in any included studies, hence do not appear in the table.
*Paper did not assess cross-cultural validity hence the quality criteria for rating the results of measurement properties (Appendix 2) were not applied to the overall measurement property result, however the translation items of COSMIN checklist were rated (Box G items 4-11).
+++ or --- strong evidence, ++ or -- moderate evidence, + or – limited evidence, ± conflicting evidence, ? indeterminate, 0 no information [+ positive, - negative rating (results)]. MHLC= Multidimensional Health Locus of Control, IHLC= Internal Health Locus of Control, PBC= Perceived Behavioural Control, PAM13= Patient Activation Measure, ENAT= Educational Needs Assessment, PEPPI-5= Perceived Efficacy in Patient–Physician Interactions 5 item Scale, SCQOA= The Stages of Change Questionnaire in Osteoarthritis, EC-17= Effective Consumer Scale, PEPPI-10= Perceived Efficacy in Patient–Physician Interactions 10 item Scale

Reliability

Adequate test-retest reliability required intraclass correlation coefficient (ICC)> 0.7 (see Appendix 2). There was limited evidence (fair rating) of inadequate test-retest reliability for the PEPPI-5 (ICC= 0.68) (180). Indeterminate evidence (poor rating) of adequate test-retest reliability was found for the EC-17 (181) (ICC= 0.71).

Measurement error

Although data for test-retest reliability can be used to calculate measurement error, only one study reported this. There was indeterminate evidence of measurement error for the PEPPI-5 (180) (limits of agreement -6.83 to 6.35) because the MIC was not defined (see Appendix 2).

Structural Validity

To demonstrate adequate structural validity, the factors identified should explain at least 50% of the variability of responses (see Appendix 2). There was strong evidence (excellent rating) that the PEPPI-5 featured an appropriate 1-factor structure (180). There was moderate evidence (good rating) that the PEPPI-10 demonstrated a two factor structure (182). There was limited evidence (fair rating) of positive structural validity for the PAM (175) and limited evidence (fair rating) that the factor structure of the SCQOA did not explain 50% of the variance (176) . There was also limited evidence (fair rating) of a negative result for structural validity of the ENAT (179). The level of evidence for the structural validity of the EC-17, MHLC and PBC (177, 178, 181) was indeterminate (poor rating).

Hypothesis Testing

The demonstration of adequate construct validity through hypothesis testing required that specific hypotheses were formulated *a priori* AND at least 75% of the results were in accordance with these (61). There was limited evidence (fair rating) for adequate construct validity for the PEPPI-5 (180) which was evaluated against; General Self Efficacy scale, Arthritis Impact Measurement Scales 2 Family and Friends scale, Short Form 36 mental component summary score, and pain numerical rating score. The EC-17 was compared with the same instruments as the PEPPI-5, however there was indeterminate evidence (poor rating) for the hypotheses tested (see Table 6.6) (181) . The study assessing PEPPI-10 did not formulate *a priori* hypotheses therefore the evidence for hypotheses testing was indeterminate (182).

Cross-cultural Validity

Cross-cultural validity is established following specified translation procedures, then comparison of two cohorts differing only in language/cultural background to test if the translated instrument accurately reflects the measurements made in the original (174). There was limited evidence (fair rating) for adequate translation of the English PAM(183) into Korean (175). The Korean PAM was not compared with the English version. There was indeterminate evidence (poor rating) for the translation of the English EC-17 (184) into Dutch (181) and no formal cross-cultural validation. There was limited evidence (fair rating) of adequate translation of the English PEPPI-10 (185) into Chinese (182) with no cross-

cultural validation. Cross-cultural comparisons were not made for the ENAT because the structural validity was inadequate in the OA group (179).

Floor and ceiling effects

Floor and ceiling effect results were rated using the quality criteria for rating the results of measurement properties in Appendix 2. There was strong evidence of absence of floor and ceiling effects for the PEPPI-5 (180), limited evidence of a ceiling effect for the PEPPI-10 (182) and indeterminate evidence for floor and ceiling effect for the EC-17 (181).

Best evidence synthesis

The instrument with the most promising level of evidence for the measurement properties available was the PEPPI-5. Of note is that these results are applicable only to the Dutch language version of the PEPPI-5. There was strong evidence for internal consistency, structural validity, and lack of floor/ceiling effects, however there was limited positive evidence for construct validity (hypothesis testing) and limited evidence of negative findings for test-retest reliability (Tables 6.6 and 6.7). There was indeterminate evidence for measurement error and no information for content validity, or responsiveness.

6.5. Discussion

Osteoarthritis self-management programs are not curative but aim to equip participants with the tools to manage their disease. It is important to measure the changes in attitudes towards

and/or capabilities regarding OA self-management to determine whether participants achieve this aim and to demonstrate the efficacy of programs. Further, it may be possible to predict the outcomes of participants by measuring attitudes towards and/or capabilities in regard to OA self-management at baseline. This may provide a basis on which to appropriately allocate healthcare resources to those that will likely benefit from such a program. Participants reporting a positive attitude toward self-management and good self-management capabilities may be prioritised for immediate engagement in a program. Conversely, individuals reporting poorer attitudes and capabilities may be targeted for supplementary therapies such as motivational coaching to improve the likelihood of successful participation in such a program. In order to test whether this is possible, we first need to identify a suitable instrument measuring attitudes towards and/or abilities regarding self-management of OA that demonstrates good measurement properties.

This systematic review is the first to synthesize the measurement property evidence for instruments assessing attitudes towards and/or capabilities regarding self-management of OA. There were a very small number of studies identified; only eight studies reported measurement properties of such instruments, each for a separate instrument. The scope of measurement properties assessed by the included studies was very limited. Internal consistency and structural validity were estimated for all instruments. Test-retest reliability (180, 181), and hypothesis testing (180, 181) were each assessed for two instruments, cross-cultural validity was addressed in three studies (175, 181, 182). Measurement error was

reported in one study (180), responsiveness and content validity were not evaluated for any of the instruments.

Given the limited measurement property evidence for the included instruments, we cannot provide a definitive, evidence-based recommendation for a particular instrument to measure attitudes towards and capabilities regarding OA self-management on the basis of good measurement properties. On balance, the instrument with the "best" measurement properties was the Dutch version of the PEPPI-5 (180). There was strong evidence that the PEPPI-5 satisfied requirements for internal consistency and structural validity. There was limited evidence for the hypotheses specified comparing PEPPI-5 scores against several other PROMs. The test-retest reliability findings were sub-optimal (i.e. ICC<0.7) which has implications regarding the standard error of the measure. Greater standard error may require larger change scores to represent 'real' change (vs error inherent in the measure) between groups over time. The evidence for measurement error of the PEPPI-5 was indeterminate because the MIC was not provided. Measurement property evidence for content validity and responsiveness of the PEPPI-5 remains unknown. The remaining instruments identified in the review demonstrated moderate evidence of positive measurement properties at best.

The PEPPI-5 was originally developed in a sample of "older people" with mixed medical diagnoses; measurement property results for internal consistency, structural and construct validity were reported for this population (185). Given the PEPPI-5 was developed for a different group of patients it may be that it has limited content validity for OA. The PEPPI-5

measures self-efficacy in obtaining both medical information and attention to chief health concern from a physician hence includes limited aspects of a patient's ability to self-manage OA. Although effective communication with a physician is important, it may not be a key outcome used to indicate the efficacy of such programs. OA self-management programs are often multidisciplinary, with input from a team of health professionals including physiotherapists, dietitians and occupational therapists (39), and some programs do not include a medical physician (186). Hence, there is a clear need to develop tools that have adequate content validity for participants of OA self-management programs.

A previous systematic review synthesized the measurement property evidence for instruments measuring self-efficacy in participants with rheumatic conditions (187). Self-efficacy is defined as the confidence that one possesses the ability to influence events that affect aspects of one's life (188). Self-efficacy is potentially an important aspect of self-management, however additional constructs may be considered such as how motivated or activated participants are to self-manage (183), or beliefs about who controls their health (177).

The previous review included participants of mixed disease groups with different rheumatic conditions (187). Given that measurement property evidence is specific to the population studied, these measurement property results cannot be extrapolated to the OA population. The population-specific nature of measurement properties also placed limitations on the studies available for this current review. Often studies were excluded at the full-text stage

because they comprised mixed disease cohorts and did not report the OA participant results separately. This limited the number of studies included.

The methodologies of the included studies were limited to the investigation of a small range of measurement properties. Internal consistency and structural validity were reported for all studies. This is a similar finding to the previous systematic review of self-efficacy in patients with rheumatic conditions (187). Although these are valuable measurement properties to establish, many measurement properties remain untested in the instruments of our systematic review. Test-retest reliability estimates the relative consistency of a measure in otherwise stable patients so that when any change is detected by the instrument, it can be attributed to the intervention rather than from measurement error of the instrument. Unfortunately, the test-retest reliability and measurement error for the included instruments are yet to be established in OA patients. Test-retest reliability was tested in a larger proportion of studies included in the systematic review on rheumatic conditions, however, the quality of the evidence was generally poor and measurement error was unreported (187). Hypothesis testing is a further property that was neglected by the majority of studies in our review. Hypothesis testing establishes whether an instrument measures the intended construct by testing the internal relationships with scores of other instruments measuring similar or different constructs (58). There is much need for future studies evaluating test-retest reliability, measurement error and construct validity of instruments measuring OA self-management attitudes and capabilities.

Cross-cultural validation was attempted in three studies that translated questionnaires; however, true cross-cultural validation comparing language versions was not conducted. This was also found in the previous review of instruments measuring self-efficacy (187). We found no evidence pertaining to content validity, responsiveness, or MID/MIC. Similar to previous conclusions (187), the recommendations arising from the present review are limited due to the small number of studies, their poor methodology, and the limited scope of measurement properties assessed. Further studies concerned with all measurement properties of existing instruments assessing self-management of OA is the only way to remedy this situation.

Some existing instruments measuring attitudes towards and/or capabilities regarding OA self-management were not featured in the systematic review because there was no measurement property evidence available. The Health Education Impact Questionnaire (heiQ) (189) evaluates the efficacy of patient education programs and has been used to evaluate OA self-management programs (15, 190). Also, the Arthritis Self Efficacy Score (ASES) measures patients' perceived self-efficacy to cope with the symptoms and limitations attributed to chronic arthritis (191) and is a published outcome of existing OA self-management programs (130, 192). The measurement properties of the heiQ and ASES remain untested in the OA population. Given the current popularity of these instruments, the measurement properties of heiQ and ASES are an important area of future research.

There were possible limitations of this systematic review; the inclusion criteria requiring studies to be published as original articles may have introduced publication bias. Unpublished

studies may have been more likely to contain evidence of negative results about measurement properties of the instruments under study. However, the inclusion of only peer-reviewed articles likely enhanced the quality of included studies, given the basic level of scrutiny required to publish. This may have improved the quality of the review rather than biasing it. While excluding non-English language studies may have introduced bias, no such studies were identified by the comprehensive search strategy.

6.6. Conclusion

This review highlights the paucity of evidence available for the measurement properties of instruments assessing attitudes towards and/or capabilities regarding OA self-management. There were many gaps in the measurement property evidence for the instruments identified. The instrument with the "best" properties assessed self-efficacy in communication with a physician; a very discrete aspect of self-management. Therefore, we were unable to make recommendations concerning instruments to assess attitudes toward and/or capabilities regarding OA self-management. Further well-designed studies of measurement properties of available instruments are required. This review may provide a starting point for researchers to identify the instruments that are currently used for this purpose in the OA population and the evidence for measurement properties available. Once we are able to identify instruments with adequate measurement properties for use in this population, we will be able to better compare the efficacy of different OA self-management programs and inform best practice for care of our patients.

Acknowledgements: The authors wish to thank Jeremy Cullis, Clinical Librarian at Macquarie University, Sydney, Australia. Jeremy generously contributed his expertise as a research librarian to assist in building the comprehensive search strategy and assisted greatly in translating the measurement property filter into the language of the different databases.

Author contributions: JPE, KM, and DH conceived the study, JPE, KM, DH, FD, NC, SM and BRL contributed to the study design. JPE and KM developed the search strategy and performed the literature search, JPE, SM and KM screened the abstracts for eligibility. JPE, SM, NC, KM and FD performed and contributed to the quality ratings. JPE, SM and KM extracted data. JE wrote the manuscript, KM, DH, FD, NC and BRL edited the manuscript. All authors read and approved the final manuscript.

Role of the funding source: There was no funding source for this study

Conflict of interest: There are no conflicts of interest to declare

CHAPTER SEVEN

This chapter contains the following paper submitted to BMC Health and Quality of Life Outcomes: Eyles, JP; Ferreira, M; Mills, K; Lucas, BR; Robbins, SR; Williams; M, Lee, H; Appleton, S; Hunter, DJ. Is the Patient Activation Measure a valid measure of osteoarthritis self-management attitudes and capabilities? Results of a Rasch analysis

Chapter 7

Is the Patient Activation Measure a valid measure of osteoarthritis self-management attitudes and capabilities? Results of a Rasch analysis

Authors

Eyles, JP[1,2], Ferreira, M[1,2] Mills, K[4], Lucas, BR[1,3] Robbins, SR [1,2] Williams, M[3], Lee, H[5] Appleton, S[6], Hunter, DJ[1,2].

Affiliations

[1]Kolling Institute of Medical Research, Institute of Bone and Joint Research, University of Sydney, Australia. [2]Department of Rheumatology, Royal North Shore Hospital and Northern Clinical School, University of Sydney, Australia. [3]Physiotherapy Department, Royal North Shore Hospital, Sydney, Australia. [4]Faculty of Medicine and Health Sciences, Macquarie University, Sydney, Australia. [5]Rehabilitation Department, Hunters Hill Private Hospital, Sydney, Australia. [6]Physiotherapy Department, Mount Wilga Private Hospital, Sydney, Australia.

Corresponding Author

Jillian Eyles

Statement from co-authors confirming authorship contribution of the PhD candidate

The co-authors of the paper: "**Is the Patient Activation Measure a valid measure of osteoarthritis self-management attitudes and capabilities? Results of a Rasch analysis**", confirm that Jillian Eyles has made the following contributions:

- Conception and design of the research
- Analysis and interpretation of the findings
- Writing of the manuscript and critical appraisal of the content

In addition to the statements above, in cases where I am not the corresponding author of a published item, permission to include the published material has been granted by the corresponding author.

Jillian Eyles Date: 1st March 2019

As supervisor for the candidature upon which this Thesis is based, I can confirm that the authorship attribution statements above are correct.

David Hunter Date: 1st March 2019

Chapter Seven: Is the Patient Activation Measure a valid measure of osteoarthritis self-management attitudes and capabilities? Results of a Rasch analysis

7.1. Abstract:

Objective: The Patient Activation Measure (PAM-13) was developed using Rasch analysis to assess knowledge, skills, and confidence in the management of one's health. Previous studies report positive relationships between PAM-13 scores, self-management behaviours and longitudinal health outcomes in adults with chronic disease. There is little extant measurement property evidence for the use of PAM-13 in specific osteoarthritis (OA) populations. This study tested measurement properties of the PAM-13 in people living with hip and knee OA.

Methods: Item response analysis demonstrated data quality. Rasch analysis evaluated the fit of the PAM-13 data to the Rasch model. Model-data fit was evaluated using infit and outfit statistics Person/item reliability and person separation indices were computed. Unidimensionality was evaluated using Principal Components Analysis of Rasch residuals and the data were assessed for item redundancy. Differential Item Functioning (DIF) examined bias in respondent subgroups and correlations tested relationships between PAM-13 and other patient-reported outcomes.

Results: Two-hundred-and-seventeen PAM-13 surveys were completed, there were no missing responses, floor or ceiling effects. Person and item reliability were acceptable (0.98 and 0.87 respectively) with good separation (person separation index 2.58).

Unidimensionality was evaluated, with 49.4% of the variance explained. There was evidence of potential local response-dependence. The Rasch fit statistics were acceptable (except for item-2). There were some issues identified with targeting of the PAM-13 items to people with higher ability and the item difficulty order was different to that proposed in original American cohorts. Significant DIF was identified for sex and educational level for a small number of items. PAM-13 scores were moderately correlated with depressive symptoms on the Depression Anxiety Stress Scale and Assessment of Quality of Life-6D. There were small correlations between PAM-13 and Knee injury and Osteoarthritis Outcome Score pain and activities of daily living (ADL) scores.

Conclusions: This study provides some evidence of adequate person and item reliability, unidimensionality, and construct validity to support the use of PAM-13 to measure patient activation in people living with hip and knee OA. Possible limitations regarding targeting, different item difficulty order, DIF and local response dependence should be investigated in future research.

7.2. Background:

Osteoarthritis (OA) is a prevalent, painful condition and a leading cause of global disability (147). As a costly (9), chronic, incurable disease, self-management interventions are recommended for the management of osteoarthritis (OA) (13). Two systematic reviews have evaluated the effects of self-management interventions that included OA patients. The first demonstrated evidence of small to moderate effects in terms of pain and functional improvements conferred by arthritis self-management interventions (14). The second was concerned specifically with OA self-management education programs and found no or small benefits from these programs (15). These reviews highlight that measures of pain and function are the most common primary outcomes for self-management interventions (14, 15). Whilst pain and function are obviously important to this population, there is a disparity between the aims of self-management programs and the outcomes used to assess efficacy. A more meaningful measurement of program efficacy would be to measure OA self-management attitudes and capabilities (15), which have been recognised as comparatively neglected domains (193).

The measurement of OA self-management attitudes and capabilities requires validated instruments that have demonstrated adequate measurement properties in populations with OA (56). Measurement properties refer to the ability of the instrument to accurately and comprehensively measure the specified construct (37) (e.g. internal consistency, reliability, validity). A recent systematic review of instruments assessing OA self-management attitudes

and capabilities found that there was very little measurement property evidence available and that further research was needed to fill this knowledge gap (41).

An instrument identified in the review was the Patient Activation Measure (PAM-13); a patient- reported outcome assessing knowledge, skill and confidence in the management of one's health (42). The measurement properties of the PAM-13 have been studied in populations with varying chronic conditions including: mental illness (194); neurological disorders (195) and multimorbidity (196, 197). Two previous studies investigated measurement properties of the PAM-13 in OA populations. The first translated the PAM-13 into Korean and provided some evidence of adequate internal consistency and structural validity (175). The second examined the responsiveness of PAM-13 in a sample of people with "arthritis", not specifically OA (198). This study aims build and extend on this prior research to provide more comprehensive evidence of measurement properties of the PAM-13 in people living with OA.

Several large cohort studies report that higher levels of patient activation measured by the PAM-13 predict better self-management behaviours and longitudinal health outcomes in adults with chronic disease (199-201). This considered, it may be possible to predict patient outcomes following OA management programs using PAM-13 scores. This would enable the identification of people likely to experience a positive treatment effect. These people could then be prioritised for participation in these programs. Conversely, people reporting poorer self-management attitudes and capabilities may be identified and targeted for supplementary

therapies (e.g. motivational coaching). Further, the efficacy of OA management programs could be measured in terms of change in patient activation. Before these potential uses of the PAM-13 are tested, it is important to establish that the measurement properties are acceptable in the OA population.

The PAM-13 developers used Rasch analysis to construct the instrument according to the Rasch measurement model (42). The Rasch model determines the measurement requirements for the construction of interval level measurement scales (202). A major advantage of using instruments developed using Rasch analysis is that the measurements can be assumed to produce interval level variables, hence, statistical tests requiring interval level variables can be used to report the results of clinical studies (203). Rasch analysis also provides a unified measurement approach to test the validity of an instrument developed using this method when it is tested in a different population of patients (56). This study had the following aims:

i) To test the measurement properties (including reliability (internal consistency), unidimensionality (structural validity) and construct validity and floor/ceiling effects) of the PAM-13 in people with hip and knee OA.

ii) To examine the relationships between PAM-13 scores and psychological, quality of life and disease-specific outcomes.

7.3. Methods:

7.3.1. Participants

This cohort study comprised participants of the Osteoarthritis Chronic Care Program (OACCP). Participants were recruited directly from Royal North Shore, Ryde (major teaching hospitals), Hunter's Hill Private and Mount Wilga (private metropolitan hospitals) hospitals in Australia via referral from rheumatologists, orthopaedic surgeons and general practitioners or joint arthroplasty waiting lists. People with symptomatic and radiographic hip and knee OA were eligible if they reported pain in the affected knee/hip on most days of the past month. Details of the program are published elsewhere (39).

Ethical approval for this study in accordance with the Declaration of Helsinki was provided by Human Research Ethics Committees:
NSPHEC 2016-LNR-007; NSPHEC 2017-LNR-005 and LNRl16/HAWKE/14. Participants provided written consent to take part in this study.

7.3.2. Data

All data were collected at the baseline assessment of OACCP as part of the normal clinical pathway. Signal joint (the predominant site of OA) was determined by clinical and radiographic examination. Anthropometric measurements were undertaken using a standardised protocol (47). Participants rated their average pain on the day of assessment using a Numeric Rating Scale (0 indicated no pain and 10 the most pain imaginable) (157). Patient-reported outcomes were collected electronically as described below.

Patient Activation Measure-13: Participants rated their level of agreement with 13 statements (Table 7.1) using a 4-point Likert scale: Totally Disagree, Disagree, Agree, Totally Agree and Not Applicable (N/A). This outcome assumes that Item-1 is the easiest to endorse, and each subsequent item is more difficult to endorse than the one before (42). The response (range 1–4) to the items are added to calculate a raw score. Responses of "not applicable" (N/A) are treated as missing. A continuous activation score is computed from the raw score using an empirically derived calibration table by Insignia Health (after January 2014). Total scores range from 0 (no activation) to 100 (high activation) (42). PAM-13 score thresholds are used to assign four stages of activation in order of ascending activation: 1. "Believes active role is important"; 2. "Confidence and knowledge to take action"; 3. "Taking action"; 4. "Staying the course under stress" (42).

Table 7.1 PAM-13 items and mean response scores

PAM-13 items	N	Mean[#] (SD)
1. When all is said and done, I am the person who is responsible for taking care of my health	217	3.4 (0.73)
2. Taking an active role in my own health care is the most important thing that affects my health	217	3.4 (0.77)
3. I am confident I can help prevent or reduce problems associated with my health	217	3.3 (0.77)
4. I know what each of my prescribed medications do	200	3.3 (0.70)
5. I am confident that I can tell whether I need to go to the doctor or whether I can take care of a health problem myself	215	3.2 (0.70)
6. I am confident that I can tell a doctor concerns I have even when he or she does not ask	217	3.3 (0.65)
7. I am confident that I can follow through on medical treatments I may need to do at home	216	3.2 (0.67)
8. I understand my health problems and what causes them	215	3.1 (0.71)
9. I know what treatments are available for my health problems	212	2.8 (0.75)
10. I have been able to maintain (keep up with) lifestyle changes, like eating right or exercising	210	2.8 (0.70)
11. I know how to prevent problems with my health	208	2.6 (0.70)
12. I am confident I can figure out solutions when new problems arise with my health	212	2.8 (0.70)
13. I am confident that I can maintain lifestyle changes, like eating right and exercising, even during times of stress	216	2.8 (0.70)

[#]PAM items are scored using “Totally Disagree”=1, “Disagree”=2, “Agree”=3, “Totally Agree”=4

The Hip disability and Osteoarthritis Outcome Score (HOOS) and Knee injury and

Osteoarthritis Outcome Score (KOOS): The HOOS (135) and KOOS (158) are disease-specific measures that have been validated in people with OA. Participants rate their symptoms, stiffness, pain, physical function, recreational activities and quality of life on a 5-point Likert scale (0- 4). The responses for the six subscales are summed and transformed to comprise six independent subscores; lower scores indicate worse problems.

The Depression, Anxiety and Stress Scale (DASS-21): Participants rate their level of agreement with 21 statements using a 4-point Likert scale (0-3). The DASS-21 subscores indicate the presence/absence of symptoms of depression, anxiety and stress (159). Higher scores indicate worse symptoms.

Assessment of Quality of Life (AQoL-6D): Participants respond to questions or statements rated using four-, five- or six-point scales. Six dimensions are reported separately including: independent living, relationships, mental health, coping, pain and senses which are combined for a standardised AQoL index. Higher scores indicate a worse quality of life (204).

7.3.3. Statistical analysis

Descriptive statistics and correlations were processed using SPSS (Version 22.0, Armonk NY: IBM Corp, USA) software. The PAM-13 responses were compared to the Rasch model

with Rasch analysis (202) using Winsteps (version 4.0.1 Linacre, J. M. (2017) Winsteps® Rasch measurement computer program. Beaverton, Oregon: Winsteps.com).

Item response frequency analysis:

Item response analysis was conducted to demonstrate data quality (174). The frequencies of each response option and missing responses were reported for each item. Floor and ceiling effects were confirmed if ≥15% of respondents answered "totally disagree" or "totally agree" to all items respectively (61).

Rasch Model Overview:

The PAM-13 was originally developed using Rasch analysis (42). A Rasch analysis compares individual items or responses of a patient reported outcome measure with a Rasch model (RM) (203). Comparison to a Rasch Model provides insight into whether scores obtained for individual items of the outcome measure can be added together to create an overall score. More specifically, it assists in determining whether the outcome measure possesses the properties of an interval scale or whether each item is stand-alone (203).

The RM assumes that responses to the items of an outcome scale are affected by the ability of the person and the difficulty of the item (62). In Rasch analysis, metrics are calculated to determine whether the relationships between ability of the person and the difficulty of the item in the study data are consistent with what would be expected to fit the RM and that the assumptions of the RM are met. 'Person ability' is calculated using the number of items of the

instrument that a person agreed with. 'Item difficulty' is estimated using the number of persons in the sample who agreed with an item. The relationship between person ability and item difficulty is clearly depicted on a person-item map. Measures of fit are used to assess whether the instrument conforms to RM requirements; infit and outfit statistics are used to indicate how accurately or predictably data fit the model (63). There is not complete agreement about the influence of sample size on fit statistics however, a sample of 200 participants has been recommended (205). For this study we aimed to recruit 250 to account for 20% non-completion rate.

Reliability and separation:

In Rasch analysis the *person reliability index* estimates the probability that the ordering of persons (based on their abilities) is preserved when they respond to further items measuring the same construct. The *Item reliability index* indicates the probability that the order of the items (based on difficulty) would be the same if the same construct was measured in a similar but discrete sample of people (62). The *person separation index* tests if the instrument is sensitive enough to distinguish between people with high and low abilities. Thresholds for acceptable indices were set at >0.8 for item reliability, >0.8 for person reliability and >2 for the person separation index (61, 63). The person-item map was used as a pictorial representation of how well the difficulty of the items aligned with the abilities of the persons who completed the survey. The alignment between item difficulty and person ability is referred to as 'targeting' (62). In addition to the Rasch analysis, internal consistency was

estimated using Cronbach's Alpha from Classical Test Theory (CTT). The threshold for Cronbach's alpha was set at 0.8.

Rasch model fit analysis:

The partial credit model was used to examine model-data fit; it was chosen because the PAM13 had four response options demonstrating different patterns of usage (62). Point-measure correlations were estimated to determine whether item responses aligned with person abilities. Point measure correlations >0.5 were considered acceptable. Infit and outfit statistics (expressed in mean square (MnSq)) indicated how well the data fit the RM. Values between 0.5 and 1.5MnSq were considered acceptable (62). An approximate global log-likelihood chi-squared statistic for overall goodness of fit was computed to indicate if the misfit of the data was large enough to be problematic (63).

Instrument performance improvement:

Rasch analysis can be used to identify overlapping items measuring similar aspects of the construct and/or items that do not fit the model well; termed *item redundancy*. Fit statistics (MnSq values) indicated whether an item might be redundant and considered for removal from the model (62). Overlapping items were also identified using the Rasch person-item map as those occupying the same location on the map. To confirm item redundancy as identified using fit statistics and/or the person-item map, it was also necessary to assess whether the content of the item overlapped with any aspect of another item. If two or more

items were similar in content, this might indicate redundancy. Following item removal, fit statistics and person-item maps confirmed whether model fit was improved.

Unidimensionality:

In Rasch analysis, structural validity is determined by confirming the unidimensionality of the construct (174). Winsteps uses a Principal Components Analysis (PCA) to create potential secondary dimensions (termed contrasts) based on the unexplained variance of the residuals, measured in eigenvalue units. The Winsteps PCA of residuals is not interpreted in the same way as Factor Analysis (FA) of the original data in CTT. For this analysis, the threshold for good evidence of unidimensionality was provided by an eigenvalue of less than 2.0 on the first contrast; (larger eigenvalues indicated the need for further investigation)(63). Where eigenvalues exceeded 2.0, a CTT factor analysis of the original data (FA) was used to evaluate unidimensionality further. The Kaiser-Meyer-Olkin measure tested sampling adequacy and Bartlett's Test of Sphericity was used to detect the presence of multiple factors.

An important assumption of the RM is that there is no local response dependency. Local response dependency can occur when items are related to each other in a way that is outside the latent trait the outcome scale is measuring (206). Local response dependency was evaluated through the calculation of Yens Q3 statistics. It is commonly recommended that these values do not exceed r= 0.7 (63). Christensen et al. (2017) proposed that a single critical threshold for Q3 statistics was not appropriate for all situations and that a Q3 value of 0.2

above the average correlation was appropriate (63). Local response dependency was assessed using both thresholds.

Differential item functioning (DIF):

DIF tested whether subgroups responded differently to items of the instrument compared with the rest of the sample. There is evidence of DIF when an item's difficulty estimate location on the latent trait varies between subgroups by more than the modelled error (62). There are two types of DIF. Uniform DIF provides information about whether the outcome scale performs similarly in subgroups while the item difficulties and person measures are held constant. Non-uniform DIF tests the performance of the outcome scale across subgroups at different levels of ability. To evaluate DIF Winsteps uses the Mantel Chi-Squared testwith (log-)odds estimates of DIF size and tests significance from a comparison of the two groups (63). DIF that exceeds 0.64 logits is considered to be moderate to large (63). The following demographic variables were used for DIF testing: gender, highest educational level (secondary vs tertiary) and signal joint (hip vs knee).

Construct validity:

Previous studies in different populations indicated PAM-13 scores were associated with the presence of depressive symptoms and health-related quality of life (207-209); hence we expected moderate correlations between DASS and AQoL scores with PAM-13 ($r>0.3$). We hypothesized that weak correlations (if any) would be observed between PAM-13 and HOOS/KOOS 'Pain' and 'Function in daily living' subscale scores ($r<0.2$). Pearson's

correlations were used for normally distributed variables, Spearman's correlations for those that were non-parametric. The thresholds for correlation size were defined as the following: 0.5 was large, 0.3 moderate, and 0.1 small (210).

7.4. Results:

Study population

Out of the 238 participants consecutively enrolled in the OACCP February 2016 to June 2017 and approached to take part in the study, 21 participants declined to participate. The characteristics of the participants who completed the PAM-13 are summarised in Table 7.2. The group excluded based on non-completion was not large enough to make statistical comparisons.

Item response frequency analysis

Of 217 attempted PAM-13 surveys, there were no missing responses, however, the N/A responses were not included in the scoring and were treated as missing data (211). The distribution of responses to the questions is depicted in Figure 7.1. The questions most commonly responded to with N/A were PAM-13 item-4 (I know what each of my prescribed

Table 7.2 Participant characteristics

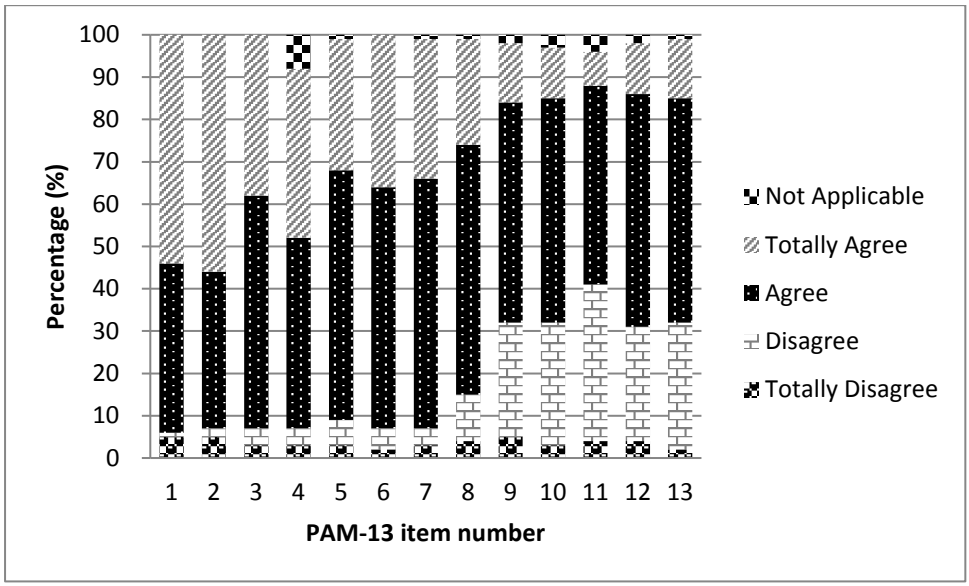
Characteristics	Included n=217	Excluded n=21
Sex female, n (%)	148 (68)	10 (48)
Age years, (SD)	65.5 (10.8)	68.7 (10.3)
Signal joint knee n (%)	183 (84)	15 (71)
Body mass index kg/m^2(SD)	30.3 (6.1)	30.0 (5.1)
PAM raw score		
PAM score mean (SD)$^{\infty}$	60.5 (11.0)	
PAM level 1 n (%)	10 (4.6)	
PAM level 2 n (%)	47 (21.7)	
PAM level 3 n (%)	123 (56.7)	
PAM level 4 n (%)	28 (12.9)	
Highest level of education		
Year 10 or equivalent n (%)	57 (26)	4 (19)
Year 12 or equivalent n (%)	28 (13)	1 (5)
Graduate degree n (%)	104 (48)	11 (52)
Post graduate degree n (%)	22 (10)	1 (5)
Missing n (%)	6 (3)	4 (19)
Work status		
Home duties n (%)	6 (3)	0
Full time n (%)	56 (26)	3 (14)
Part-Time n (%)	21 (10)	3 (14)
Retired n (%)	100 (46)	8 (38)
Volunteer n (%)	4 (2)	1 (5)
Other n (%)	25 (12)	2 (10)
Missing n (%)	5 (1)	4 (19)
Private hospital n (%)	160 (74)	21 (100)
Public hospital n (%)	57 (26)	0
Average pain in last week on VAS$^{\alpha}$	n = 214	n = 17
VAS mean (SD)	4.0 (2.3)	3.4 (2.5)
KOOS*	n = 179	n = 12
Pain mean (SD)	52.3 (17.7)	53.6 (14.1)
Function in daily living mean (SD)	58.0 (19.8)	59.3 (19.1)

HOOS^	n = 32	n = 3
Pain mean (SD)	57.7 (19.2)	61.7 (16.7)
Function in daily living mean (SD)	59.5 (18.1)	48.3 (20.2)
DASS-21±	n = 214	n = 16
Depression mean (SD)	7.2 (8.66)	5.2 (6.8)
Anxiety mean (SD)	5.1 (7.3)	3.8 (4.3)
Stress mean (SD)	8.8 (8.5)	7.5 (5.2)
AQoL#	n = 198	n = 6
Independent living mean (SD)	68.4 (19.3)	68.5 (15.5)
Social relationships mean (SD)	75.4 (20.0)	86.7 (10.3)
Mental health mean (SD)	69.8 (21.9)	74.0 (11.6)
Coping mean (SD)	65.3 (20.3)	69.3 (8.8)
Pain mean (SD)	45.8 (22.3)	55.0 (18.7)
Physical senses mean (SD)	81.8 (10.9)	79.7 (6.5)
AQoL summed index mean (SD)	68.4 (14.9)	72.1 (8.6)

∞PAM: Patient Activation Measure- 0= worst, 100= best, however participants with scores of 0 or 100 were excluded from having a final score. PAM level 1 = least activated, 4 = most activated.
α VAS: Visual analogue scale- average pain over the last week 0 = no pain, 10 = worst pain imaginable.
* KOOS: Knee injury and Osteoarthritis Outcome Score- 0=worst, 100=best
^ HOOS: Hip disability and Osteoarthritis Outcome Score- 0=worst, 100=best
± DASS: Depression Anxiety Stress Scales- 0= best, 42 = worst.
Assessment of Quality of Life Instrument- Standardised scores- 0=worst, 100=best.

medications do) and item-11 (I know how to prevent problems with my health), although N/A responses only comprised 2% (49/2821) of the total responses to these items. The most frequent response category overall was "agree" which comprised 1458/2821 (52%) of the total responses, followed by "totally agree" with 813/2921 (29%) responses. The "disagree" and "totally disagree" categories were much less frequent comprising 403/2821 (14%) and 98/2821 (3%) of all responses, respectively.

Figure 7.1 Responses across items of the PAM-13; agreement and not applicable categories



The mean response scores (range 1-4) for each item decreased from 3.4 (SD 0.73) for item-1 to 2.6 (SD 0.70) for item-11 (see table 7.1). Although the mean response demonstrated an overall trend of decreasing as the questions became more difficult with subsequent items, the individual item order did not follow the originally established order of the questions (42): for example, the mean for item-11 (mean 2.6, SD 0.70) was smaller than the means for item-12 and item-13 (mean 2.8, SD 0.70). Floor and ceiling effects were not detected; one percent (2/217) and three percent (7/217) of participants answered with 'totally disagree" and "total agree" to all items respectively.

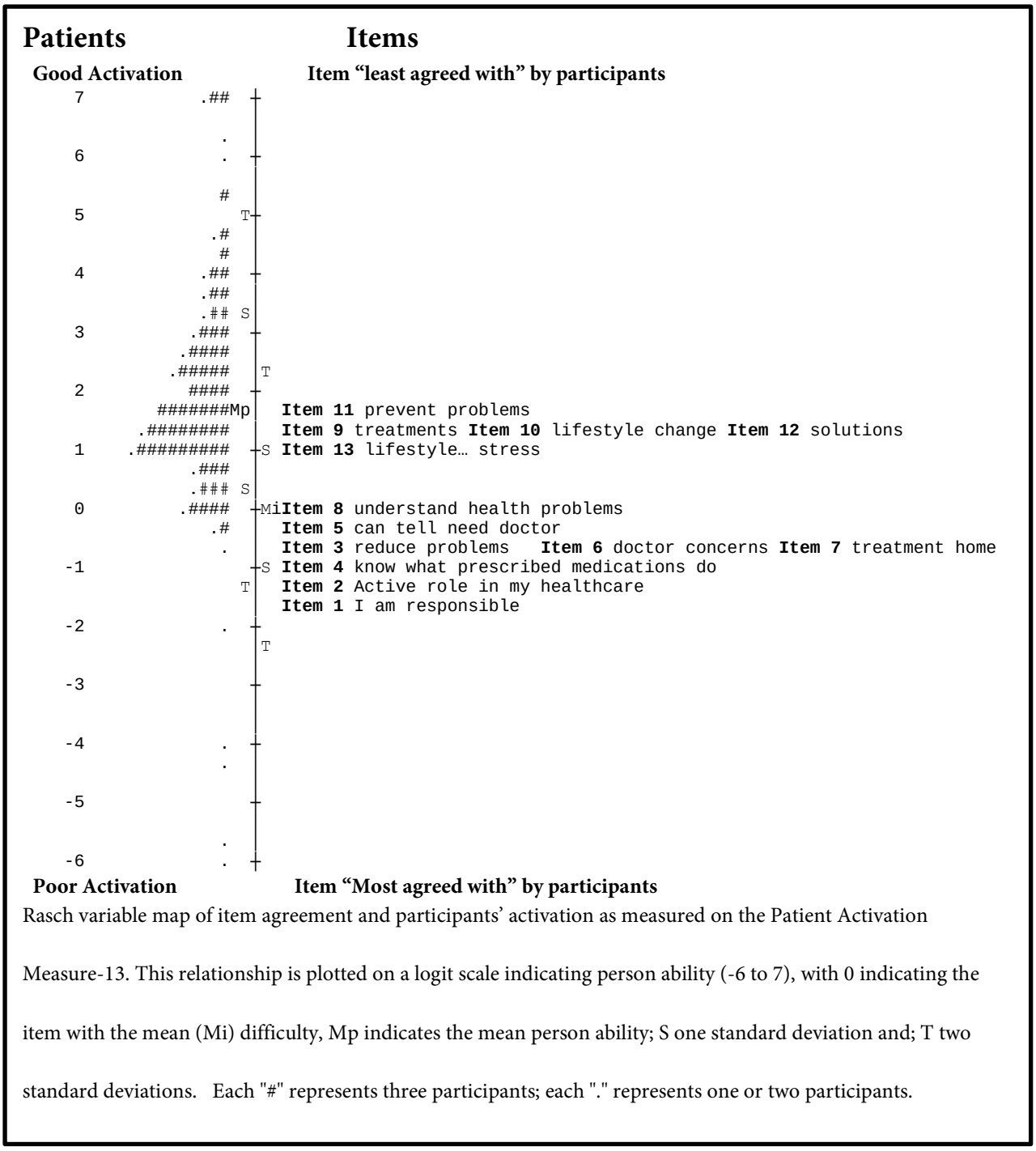
Reliability and separation

The person and item reliability of the PAM-13 was adequate as indicated by; person reliability index 0.87, item reliability index 0.98. The person separation index was 2.58 indicating good separation. In addition Cronbach's Alpha indicated adequate internal consistency; α= 0.92.

Rasch model fit analysis

There were high positive point measure correlations of r= 0.58-0.78 for all PAM-13 items. The relationship between the difficulty of the items and the ability of participants expressed in logits is depicted in Figure 7.2. Overall, the mean difficulty of the PAM-13 questions was lower than the mean ability of this sample. The mean PAM-13 item difficulty was shown at 0 logits, and the mean response of participants was almost 2 logits higher, 37% (81/217) people had abilities that exceeded the two most difficult items. Figure 7.2 also shows that the items were not evenly-spread with several items having very similar item difficulty (see items 3,6 and 7; items 9, 10 and 12). Moreover, the item difficulty did not ascend uniformly with each subsequent item. This is confirmed by item difficulty calibrations (Table 7.3) which showed item difficulty order was different to the original PAM-13. Fit statistics are summarised in Table 7.3. Items fit the RM apart from item-2 with infit and outfit statistics of 1.58 and 1.97 Msq respectively, indicating under-fit. However, the global fit statistic indicated overall adequate fit of the data to the model (log-likelihood χ^2 = 3901.0644, 3927 +- 5 degrees of freedom, P = 0.612).

Figure 7.2 Person-item map of study participants and PAM-13 items



Rasch variable map of item agreement and participants' activation as measured on the Patient Activation Measure-13. This relationship is plotted on a logit scale indicating person ability (-6 to 7), with 0 indicating the item with the mean (Mi) difficulty, Mp indicates the mean person ability; S one standard deviation and; T two standard deviations. Each "#" represents three participants; each "." represents one or two participants.

Table 7.3 Item fit statistics for the PAM-13

PAM items (in order of difficulty)	Difficulty calibration (logits)	Mean ability logits (SD)	Infit Mean squared	Outfit Mean squared
1	-1.52	-2.65 (3.26)	1.21	1.13
2	-1.50	-2.09 (3.58)	**1.58**	**1.97**
4	-1.07	1.95 (1.80)	1.27	1.16
6	-0.75	-5.17 (1.92)	0.80	0.75
3	-0.75	-4.02 (2.70)	0.95	0.89
7	-0.55	1.68 (0)	0.95	1.04
5	-0.40	1.36 (1.01)	0.85	0.80
8	0.11	1.93 (0.26)	0.74	0.70
13	1.14	1.68 (0)	1.12	1.44
9	1.17	2.54 (1.87)	0.91	0.95
10	1.19	2.34 (2.01)	0.96	0.97
12	1.23	2.54 (2.17)	0.96	0.96
11	1.68	2.55 (1.62)	0.93	1.29

Note: Results in bold indicate values that are beyond the ideal cutoffs for infit and outfit statistics (i.e. Msq of 0.5- 1.5)

Unidimensionality and structural validity

The Rasch dimension demonstrated that the persons and items within the analysis explained 49.4% of the variance (49.8% was expected if the sample fit model perfectly), with an eigenvalue of 12.70. The first contrast gave an eigenvalue of 2.5. Unidimensionality was further assessed using CTT FA. The data was adequate for the FA (Kaiser-Meyer-Olkin value = 0.88 and Bartlett's Test of Sphericity χ^2=1404.0, df 78, p <0.001). Using a scree plot and principal axis factoring, the PAM-13 loaded on one factor which explained 45.0% of the variance and suggested unidimensionality. In the assessment of local response dependence, the Yens Q3 values did not exceed the first common threshold of r=0.7 suggesting the

absence of local response-dependence. According to the second threshold, five items exceeded the Q3 value of 0.2 above the average correlation indicating the presence of local response-dependence.

Differential item functioning

No significant uniform DIF was found for people with hip OA compared with those with knee OA. There was significant uniform DIF for item-13 which was more easily endorsed by women; (DIF contrast = 0.98 logits, Mantel chi-squared statistic χ^2_M= 11.83, p = 0.001) compared to men. Item-7 was easier to endorse for people who reported their highest educational level was tertiary vs those whose highest level was high school (DIF contrast =0.85 logits, χ^2_{MH}= 4.67, p = 0.031). Conversely, people whose highest level of education was high school found Item-11 easier to endorse than those with tertiary level education (DIF contrast= 0.68 logits, χ^2_M= 6.25, p=0.012). The subgroups tested in this sample were not large enough to test for non-uniform DIF.

Instrument performance improvement

The person ability and item responses were assessed on person-item maps that depicted the logit values for all possible response options. The person-item map in figure 2 summarises the mean logit response across all response options. Although there were overlapping and similar item difficulties for items 3, 6 and 7 (Figure 7.2 and Table 7.3), these items measured different aspects of the construct and were deemed inappropriate for removal. Similarly, items 9, 10 and 12 were overlapping (Figure 7.2), however, measured different aspects of the construct

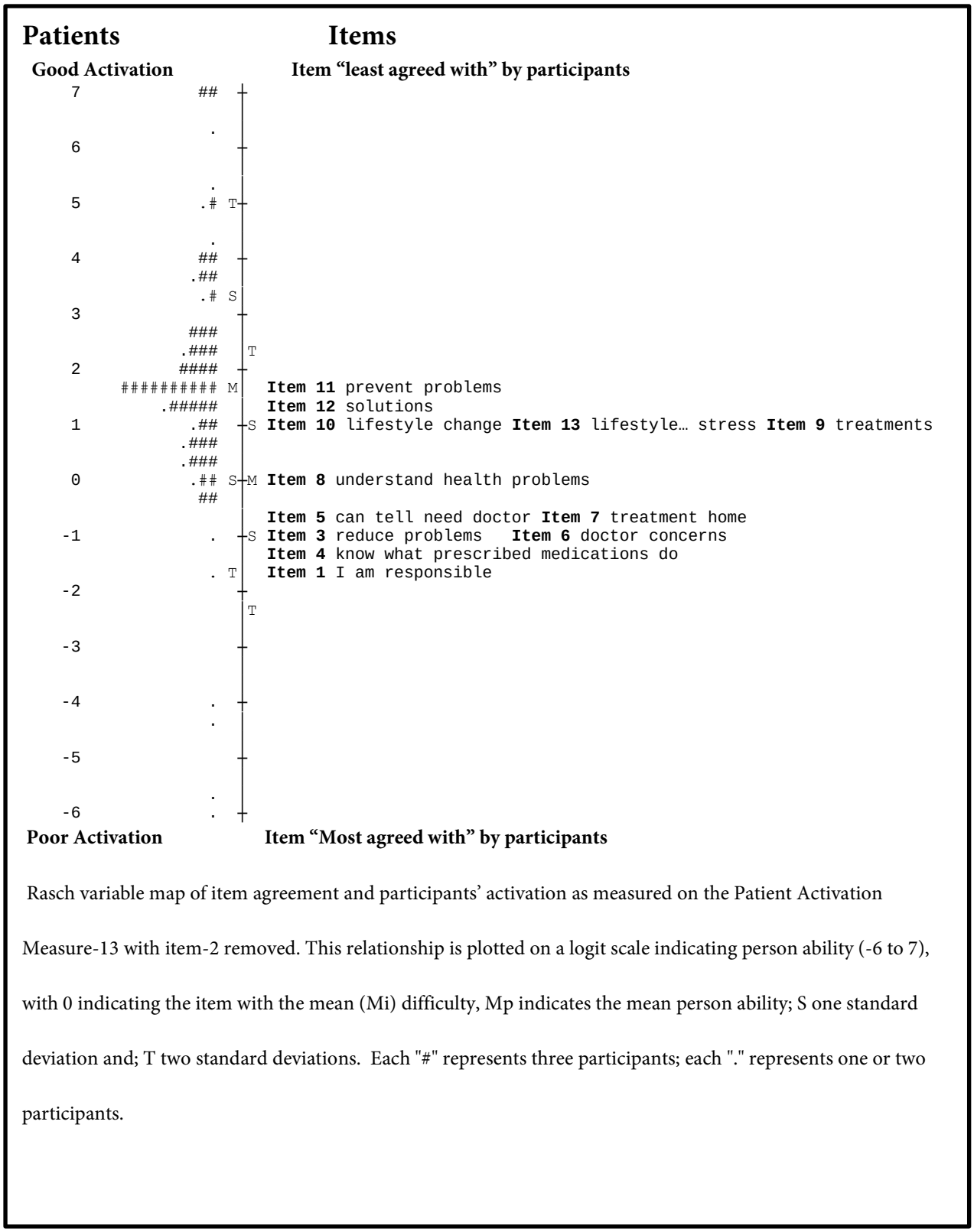
and were retained. Item-2 demonstrated poor fit statistics and was similar in item difficulty and content to item-1 (Table 7.3) so was removed. Removal of item-2 resulted in a slight improvement in the spread of the PAM-13 items (Figure 7.3). Reliability remained adequate (person and item reliability 0.87 and 0.98, respectively) and there were high positive point measure correlations (r = 0.61 to r = 0.79). The fit statistics for the revised model (Table 7.4) revealed item-1 outfit statistic 1.56 MnSq, while the remaining items were acceptable. The PCA showed 49.9% of the variance was explained by the model (compared with 50.0% expected) with an eigenvalue of 12.0. The first contrast resulted in an eigenvalue of 2.2. The analysis following removal of item-2 did not improve the performance of the instrument adequately to recommend removal of this item in this population.

Table 7.4 Item fit statistics for the PAM-13 following removal of items

PAM items (in order of difficulty)	Difficulty calibration	Mean ability logits (SD)	Infit Mean squared	Outfit Mean squared
1	-1.73	-2.45 (3.63)	1.40	**1.56**
4	-1.25	1.94 (2.07)	1.32	1.19
6	-0.92	-5.25 (2.02)	0.83	0.78
3	-0.92	-4.10 (2.77)	1.04	0.99
7	-0.72	2.80 (0)	1.00	1.11
5	-0.55	1.28 (0. 97)	0.88	0.84
8	-0.02	2.40 (0.40)	0.75	0.7
13	1.07	2.80 (0)	1.14	1.42
9	1.10	2.98 (2.37)	0.90	0.93
10	1.13	2.43 (2.04)	0.95	0.95
12	1.17	2.78 (2.15)	0.97	0.97
11	1.65	2.64 (1.62)	0.92	1.23

Note: Results in bold indicate values that are beyond the ideal cutoffs for infit and outfit statistics (i.e. Msq of 0.5- 1.5)

Figure 7.3 Person-item map of study participants and PAM-13: removal of item-2



Rasch variable map of item agreement and participants' activation as measured on the Patient Activation Measure-13 with item-2 removed. This relationship is plotted on a logit scale indicating person ability (-6 to 7), with 0 indicating the item with the mean (Mi) difficulty, Mp indicates the mean person ability; S one standard deviation and; T two standard deviations. Each "#" represents three participants; each "." represents one or two participants.

Construct validity

The correlations between the PAM-13 scores and other variables are summarized in Table 7.5. Lower activation scores were moderately correlated with the presence of depressive symptoms on the DASS; r=-0.26, (95% Confidence interval (CI) -0.38, -0.14). Higher activation scores correlated moderately with higher health-related quality of life score as measured on the AQoL; r=0.32 (95% CI 0.18, 0.47). There were small correlations between PAM-13 and KOOS pain and ADL scores (r=0.13 (95% CI 0.03, 0.29) and r=0.15 (95% CI 0.03, 0.31). There were no significant correlations between PAM-13 and HOOS pain or function scores.

Table 7.5 Correlations between Patient Activation Measure-13 scores and other variables

	Expected correlation*	Actual correlation	95% Confidence interval
DASS depression (n=205)	<-0.3	-0.26	-0.38, -0.14
AQoL (n=190)	>0.5	0.32	0.18, 0.47
KOOS Pain (n=171)	<0.2	0.13	0.03, 0.29
KOOS ADL (n=171)	<0.2	0.15	0.03, 0.31
HOOS Pain (n=31)	<0.2	-0.06	-0.47, 0.39
HOOS ADL (n=31)	<0.2	-0.23	-0.54, 0.15

Pearson's correlations were used for normally distributed variables, Spearman's correlations for those that were non-parametric. DASS: Depression Anxiety Stress Scales- 0= best, 42 = worst. AQoL: Assessment of Quality of Life Instrument- Standardised scores- 0=worst, 100=best. KOOS: Knee injury and Osteoarthritis Outcome Score- 0=worst, 100=best. HOOS: Hip disability and Osteoarthritis Outcome Score- 0=worst, 100=best

7.5. Discussion:

Adequate person and item reliability was demonstrated for the PAM-13 and unidimensionality was evaluated. There were some issues with targeting items to people with higher abilities and the item-order was different to that expected for the PAM-13. Rasch analysis revealed that Item-2 under-fit the model and its removal resulted in a very slightly improved model fit, but not enough to recommend its removal. There was evidence of a difference in item response based on sex and educational status, though this was limited to a small number of items. The presence of depressive symptoms and AQoL scores correlated moderately with PAM-13 as expected.

International studies commonly report a different item difficulty order to the original order for PAM-13 published in American cohorts (42, 175, 195, 212-214). This was consistent with the findings of our study with the exception of three items: -1 (When all is said and done, I am the person who is responsible for taking care of my health); -2 (Taking an active role in my own health care is the most important thing that affects my health) and -12 (I am confident I can figure out solutions when new problems arise with my health) (42). These were also the only items consistent with the original order in a Canadian study of participants with neurological conditions (195). Items-1 and -2 alone followed the original order of item difficulty in a study of adults in Korea living with OA. Item-1 was the 'easiest' item in a Danish study (213), but not in studies of the German and Italian PAM-13 in people with chronic conditions (212, 214). Differences in item-difficulty order seen in our study and other populations may be attributed to specific disease and cultural factors. The differences in self-management tasks required and the corresponding difficulty of these should be considered in the context of the health conditions and populations in which the PAM-13 is used.

Unidimensionality of the PAM-13 was assessed, almost 50% of the variance was explained by the items and participant responses. This percentage of explained variation was higher than reported in other disease populations (212-214), but not as high as that reported for the Korean version of PAM-13 tested in an OA sample (57.5%) (175). The limited proportion of variance explained suggests there may be other additional factors that comprise this construct that are not captured by the items of the instrument. On the other hand, it may indicate simply that the items were of similar difficulty and the participants in the study were of

similar ability (63). This study relied on the PCA of Rasch residuals and conventional factor analysis. Further information on the unidimensionality of PAM-13 in this population using other statistical tests such as confirmatory factor analysis based on polychoric correlations or further Rasch based tests may be valuable in future research to explore this further.

The assessment of local response dependence using two different thresholds of Q3 values yielded conflicting results. The conventional threshold Q3 value indicated the absence of local response dependence. The second threshold of the mean Q3 + 0.2 suggested the presence of local response dependence. To confirm these results, it would be helpful to attempt to replicate these results in future studies. This could also provide further evidence regarding the potential for different results produced by commonly recommended thresholds versus thresholds that are influenced by the characteristics of the dataset being analysed.

There were issues identified with the targeting of the PAM-13 items. Specifically, there was a lack of items of sufficient difficulty for the participants with greater ability. This could affect the precision of the measure in these people. A possible way of dealing with both the low proportion of explained variance and limited targeting of items to people with higher abilities may be to develop an OA-specific version of the PAM-13 in a similar way the version was developed for mental health (PAM-MH) (215). An important implication of modifying the PAM-13 to be condition-specific would be a loss of the ability to compare populations and the relative impact of different medical conditions and/or treatments. Further, people with

OA commonly report the presence of several chronic comorbidities (216); it is arguably more useful to use a generic instrument and consider self-management of health in general.

There were a few incidences of significant DIF in this study, however, it is important to recognise that the size of the subgroups used in the DIF analysis were low, which may have increased the type I error rate in these analyses (217). Sample sizes of 200 per subgroup have been recommended for DIF analyses (217). Significant DIF was found for item-13 suggesting women find it more difficult to endorse this item. This is consistent with one study in a different population (212) however, there was no significant DIF identified for gender in the OA Korean PAM-13 study (175). Our analysis found that PAM Item-7 was easier to endorse for those people with higher formal educational level, yet, an Italian study found the opposite, reporting that people with higher education levels found this item more difficult to endorse (212). In our study, item-11 was harder to endorse for participants with higher education levels, this was not reported in other studies.

Given the concerns that low sample size may have increased the incidence of type I error in this analysis and the lack of consistency of DIF reported for the different items of PAM-13 in other studies we did not further evaluate the DIF in this study. To further evaluate DIF for PAM-13 in this population, studies with larger subgroup sizes should be used. If DIF is found for the same items in future studies, there are several ways this could be managed such as removal of those items (218).

This is the first time, to our knowledge that a study, has examined measurement properties of the English language version PAM in a sample of people living with OA. There is growing interest in the utility of the PAM-13, particularly in the United Kingdom where the PAM-13 is being appraised as a tool used to evaluate care for chronic conditions in the National Health Service (219). It is important to improve our understanding of the measurement properties of PAM-13 in different disease populations and this study is a valuable contribution to this growing body of evidence.

There are some limitations to the applicability of this study which included a fairly homogenous population from a higher socio-demographic region of Australia. Future studies should aim to include a less geographically and socio-demographically homogenous sample. There was also a large proportion of people in our study with knee OA so that the sample was less representative of people with hip OA. Future studies should assess larger groups of participants with hip OA to ensure that accurate measurement properties are available for people with this disease.

7.6. Conclusion

There is limited extant measurement property evidence available to support the use of any instrument assessing OA self-management attitudes and capabilities. This study provides evidence of adequate person and item reliability, unidimensionality, and construct validity to support the use of PAM-13 to measure patient activation in people living with OA. Potential areas for concern regarding the PAM-13 responses from this sample include possible local

response dependence, DIF and issues with targeting. Studies in larger cohorts of people with OA is recommended to provide further information on measurement properties not just for the purposes of research, but to provide information about how the PAM-13 can be used with individual OA patients in the clinic.

Acknowledgements

We would like to acknowledge the participants of the OA chronic disease management programs at Hunters Hill, Royal North Shore and Mount Wilga Hospitals who graciously completed the PAM-13 and the other patient-reported outcomes used as part of the program.

CHAPTER EIGHT

This chapter contains the following paper to be submitted for publication in a peer-reviewed journal:

Eyles, JP; Mills, K; Lucas, BR; Robbins, SR; O'Connell, RL; Williams, M; Lee, H; Appleton, S; Hunter, DJ. Examining patient activation as a predictor of short-term outcomes following an osteoarthritis chronic disease management program.

Chapter 8

Examining patient activation as a predictor of short-term outcomes following an osteoarthritis chronic disease management program.

Authors

Eyles, JP[1,2,3]; Mills, K[4]; Lucas, BR[1,3]; Robbins, SR [1,2]; O'Connell, RL[5]; Williams, M[3]; Lee, H[6]; Appleton, S[7]; Hunter, DJ[1,2].

Affiliations

[1]Kolling Institute of Medical Research, Institute of Bone and Joint Research, The University of Sydney, Australia. [2]Department of Rheumatology, Royal North Shore Hospital and Northern Clinical School, Faculty of Health and Science, The University of Sydney, Australia. [3]Physiotherapy Department, Royal North Shore Hospital, Sydney, Australia. [4]Faculty of Medicine and Health Sciences, Macquarie University, Sydney, Australia. [5]National Health and Medical Research Council Clinical Trials Centre, The University of Sydney, Australia. [6]Rehabilitation Department, Hunters Hill Private Hospital, Sydney, Australia. [7]Physiotherapy Department, Mount Wilga Private Hospital, Sydney, Australia.

Corresponding Author

Jillian Eyles

Statement from co-authors confirming authorship contribution of the PhD candidate

The co-authors of the paper: "**Examining patient activation as a predictor of short-term outcomes following an osteoarthritis chronic disease management program.**", confirm that Jillian Eyles has made the following contributions:

- Conception and design of the research
- Analysis and interpretation of the findings
- Writing of the manuscript and critical appraisal of the content

In addition to the statements above, in cases where I am not the corresponding author of a published item, permission to include the published material has been granted by the corresponding author.

Jillian Eyles Date: 15th October 2018

As supervisor for the candidature upon which this Thesis is based, I can confirm that the authorship attribution statements above are correct.

David Hunter Date: 15th October 2018

Chapter Eight: Examining patient activation as a predictor of short-term outcomes following an osteoarthritis chronic disease management program

8.1. Abstract

Objective: To examine baseline patient activation as a predictor of short-term symptomatic outcomes and quantify the change in Patient Activation Measure (PAM-13) scores following participation in the Osteoarthritis Chronic Care Program (OACCP). We also assessed the relationship between PAM-13 and Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) longitudinal scores.

Methods: One-hundred-and-ten OACCP participants with knee osteoarthritis were followed with assessments at 0-, 12- and 26-weeks. Demographic variables (e.g. sex, age, employment status), timed-up-and-go (TUG), PAM-13, Depression Anxiety Stress Scale and WOMAC were collected. Multivariable linear regression examined the relationships between baseline PAM-13 scores and WOMAC pain and function change scores at 12- and 26-weeks.

Results: Complete 12- and 26-week data were available for 89 and 66 participants respectively: mean age 67.1 years; 75% female; 14% waitlisted for total joint arthroplasty. Baseline PAM-13 did not predict pain (β=0.10 (95%CI-0.12, 0.31) p=0.50) or function (β=0.08 (95%CI-0.11, 0.28), p=0.40) scores at 12- or 26-weeks. Employment status and TUG were independently associated with 12-week change in pain and function. No significant predictors of changes in 26-week pain or function were identified. Following 26-weeks of the

OACCP, PAM-13 scores improved by 5.8 points (n= 66, 95%CI 1.89, 9.78). PAM-13 change scores were not significantly correlated with pain or function change scores.

Conclusions: Although improvements in pain, function and PAM-13 scores were achieved following participation in the OACCP, baseline and change PAM-13 scores did not predict changes in pain and function. Employment status and TUG were independent predictors of changes in pain and function.

8.2. Introduction

Osteoarthritis (OA) is a leading cause of global disability (8). International guidelines for the management of OA recommend a combination of non-surgical, non-pharmacological and pharmacological treatments (13). Despite the availability of concordant, high-quality, evidence-based management guidelines, these have not been implemented sufficiently into clinical practice (12). Considerable discrepancy persists between recommended OA care, and the actual care that is received by patients (24, 220). To address this discrepancy, a number of OA management programs (OAMPs) have been implemented internationally (22). Some studies have reported that a substantial proportion of participants achieve clinically important improvements from these programs; however, some do not achieve such benefits (27, 29, 30). Further, some people report worsening symptoms despite their participation (40). Identifying participants most likely to benefit from OAMPs would enable clinicians to prioritise these individuals for entry to these programs. The people identified as unlikely to benefit may be referred for adjunctive therapies (e.g. motivational counselling) to improve their outcomes.

Several studies have attempted to identify predictors of response to OAMPs (30, 33, 34, 39, 130). There were two studies that identified higher log-odds of women improving than men (34, 39). Another two studies associated poorer outcomes with total joint arthroplasty (TJA) waitlist status (33, 40). Otherwise, the predictors in these studies were inconsistent. Given this inconsistent evidence, it is necessary to consider alternative explanatory variables that are plausible predictors of outcome to intervention.

Published international OAMPs consistently provide participants with: support for self-management; education; and exercise as core treatments (22). Additional optional interventions are offered by some programs including: weight loss; medication review; provision of assistive devices; psychosocial support; or orthotics (22). Participants of OAMPs are often prescribed complex multimodal treatment including lifestyle interventions (particularly weight loss and exercise). These complex interventions require a considerable commitment and substantial, sustained health behaviour change.

Given the importance of health behaviour change to the success of lifestyle interventions, it is necessary to consider the influencers of behaviour change. A scoping review with systematic searches examined barriers and facilitators to exercise participation in people with hip and knee OA (221). Some of the important barriers to exercise participation identified in the review included; lack of knowledge about exercise for OA; negative attitudes to health and exercise; and low motivation (221). Conversely, education or knowledge about OA, positive expectations of exercise effects, strong levels of motivation and determination were identified as facilitators to exercise participation (221). A systematic review of the determinants of adherence to lifestyle interventions in obese adults identified several barriers to behaviour change including: poor motivation and gaps in knowledge/lack of awareness (222). Both of these reviews suggest that knowledge, attitudes and motivation play a critical role in health behaviour change for lifestyle interventions.

The knowledge, attitudes and motivation of participants towards OA self-management and health behaviour can be regarded as a similar concept to 'patient activation'. Patient activation is described as the self-reported confidence, knowledge and skills regarding self-management of one's health (42). This construct is quantified using the Patient Activation Measure-13 (PAM-13). Previous studies report that higher scores on the PAM-13 are associated with better self-management behaviours and outcomes including: participation in regular exercise; attending to the fat content of foods; medication and physical therapy adherence; self-management knowledge; and appropriate care-seeking behaviours (199, 200, 223). Much of the existing PAM-13 evidence is derived from large North American primary care cohorts (199). As PAM-13 is now being implemented elsewhere, e.g. the United Kingdom's National Health Service, there is recognition of the need to understand the utility of PAM-13 within different contexts (219).

This study examined the relationships between PAM-13 scores and symptomatic improvements in participants of an OAMP. An important aspect to consider when investigating patient activation as a predictor is the timepoint at which most improvements are gained from OAMPs. Previous studies have demonstrated that symptomatic improvements gained in the first three to six months of OAMPs tend to decline over time (27, 130). It is also important to consider that the changes in symptoms achieved in the early months of a program may greatly influence the long-term success of participants (130). For these two reasons, the present cohort study examined PAM-13 scores as a predictor of short-term symptomatic outcomes. Our primary aim was to determine if baseline PAM-13 scores

predicted change in Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain and function scores following 12 weeks of an OAMP. Secondary outcomes included; to examine PAM-13 scores as a predictor of change in symptoms at 26 weeks, to quantify the change in 'Patient Activation' between baseline and 26 weeks and assess the relationship between PAM-13 and WOMAC change scores.

8.3. Methods

Participants and data collection

This study comprised a clinical cohort of consecutive participants with symptomatic and radiographic knee OA recruited for OAMPs at a major teaching hospital and a private metropolitan hospital in New South Wales, Australia. These OAMPs were based on the same model of care (134), the main difference being one was in a public hospital and one a private hospital setting. Both programs henceforth will be referred to as Osteoarthritis Chronic Care Program (OACCP). Participants were recruited through referral by rheumatologists, orthopaedic surgeons, general practitioners and TJA waiting lists. People with a diagnosis of knee OA were eligible if they reported pain in the affected knee on most days of the past month (39), there were no exclusion criteria. Participants provided consent to participate in this study. Ethical approval for this study in accordance with the Declaration of Helsinki was provided by Human Research Ethics Committees (HREC):

NSPHEC 2017-LNR-005 and RESP/16/11, HREC reference: LNRl16/HAWKE/14.

Participants provided consent to take part in this study and were given the opportunity to opt out of having their data included in this research.

The objectives of the OACCP were to: reduce OA pain, increase functional abilities and quality of life of participants through the provision of tailored interventions delivered by a multidisciplinary team and referral to appropriate community-based services. At initial assessment, a musculoskeletal (MSK) coordinator (experienced physical therapist/exercise physiologist) provided participants with education about their OA and associated comorbidities, set patient-oriented goals, prescribed behavioural change strategies and an individually tailored exercise program. The exercise program comprised of strength and cardiovascular training; and was evaluated and progressed at reassessments (12, 26 and 52 weeks). Participants also attended a multidisciplinary clinic for consultations with a rheumatologist, dietitian, occupational therapist, social worker or orthotist according to their individual clinical needs.

Outcome Measures

As per standard clinical practice for the OACCP, demographic data were recorded at baseline including: age; educational level (finished/did not finish high school); and employment status (engaged/ not engaged in paid employment). Standardised outcomes collected at each OACCP assessment stage included the following:

i) Measures of height and weight were performed using a standardised protocol (47).

ii) Participants rated their average pain on the day of assessment using an 11-point Numerical Rating Scale - NRS (0 no pain, 10 worst pain imaginable) (157).

iii) The Knee injury and Osteoarthritis Outcome Score (KOOS)(158) requires participants to rate their: Symptoms; Stiffness; Pain; Physical Function; Recreational Activities; and Quality of Life on 5-point Likert scales. The KOOS subsumes the WOMAC questions enabling calculation of WOMAC pain, stiffness and function subscales (137) which has demonstrated acceptable measurement properties for people living with OA (224). Higher scores indicate worse symptoms of pain, stiffness or function.

iv) The Depression, Anxiety and Stress Scale 21 item version (DASS-21) requires participants to rate how much 21 separate statements applied to them over the past week using a 4-point Likert scale (0: Did not apply to me at all – never; 3: Applied to me very much, or most of the time - almost always). DASS depression responses were added, doubled and dichotomised according to thresholds indicating symptoms of moderate depression or worse (scores ≥ 14/42) and no symptoms to mild symptoms of depression (scores < 14) (159).

vi) The Timed Up and Go (TUG) measures the time taken for participants to stand from a chair (with arms), walk in a straight line for three metres, turn around 180 degrees and walk back to the chair and sit down (142).

vi) A modified version of the Self-Administered Comorbidity Questionnaire estimates a commodity score based on participant responses to 'has your doctor told you that you have any of the following problems?', 21 commonly reported conditions were listed plus an 'other' category (141). The comorbidity count was categorised into low (0-1), moderate (2-3) and high (≥4) groups.

In addition to the standard OACCP outcomes, participants were asked to complete the PAM-13 at baseline and 26-week assessments. The PAM-13 requires participants to rate their agreement with 13 statements using a 4-point Likert scale. Responses to the items were added to calculate a raw score. A continuous activation score was calculated using an empirically derived calibration table by Insignia Health (after January 2014) ranging from 0= no activation to 100= high activation (42).

Statistical Analyses

Statistical analyses were conducted in SPSS (Version 24.0, Armonk NY: IBM Corp, USA). The distribution of variables was assessed before analysis through visual inspection of histograms and Q-Q plots. Descriptive statistics including frequencies (percentage), means and standard deviations summarised participant characteristics. Paired t-tests compared PAM-13, WOMAC pain and function scores between baseline, 12- and 26 weeks. Pearson's correlation coefficient was used to measure the strength of linear association between change scores of the PAM-13, WOMAC pain and function. The thresholds used to define correlation size were; 0.5 strong, 0.3 moderate, and 0.1 weak (210).

Potential independent predictor variables were identified *a priori* through literature review and consensus of the authors. These included: PAM-13 scores; age; sex; level of comorbidity; presence of depressive symptoms; TJR waitlist status; educational level; employment status; and TUG. First, the unadjusted analysis was performed using simple linear regression to study relationships between potential baseline predictors and WOMAC pain and function

change scores. In the second step, models included an adjustment for baseline WOMAC pain/function scores. Variables with a P-value of ≤ 0.1 were included as candidate variables in multivariable models which were built using a backward selection, forced entry technique. At each step, the least significant variable (p-value<0.05) was removed from the model and the beta coefficients were checked; if the coefficients changed by ≥ 10% the variable was retained in the model as a confounder, if not it was removed. The underlying assumptions of the linear regression model were tested (54).

Power calculations for this study were based on accommodating six predictors at 10-15 cases per predictor which is a conservative estimate for multivariable linear regression and takes into account the degrees of freedom required for the univariable pre-screening (55). Therefore we aimed to include approximately 90 participants in the analysis. With knowledge of the expected loss-to-follow-up of participants of this clinical program, we aimed to recruit 112 participants to accommodate a 20% drop-out rate.

8.4. Results

Of 117 participants with knee OA recruited to the OACCPs between February- December 2017, 110 consented to take part in the study and completed baseline PAM-13 questionnaires in addition to the standard outcomes (see Figure 8.1). The cohort was predominantly female (75%), 61.1 (Standard Deviation [SD] 10.0) years of age, overweight with mean BMI 31.5 (SD 6.2) kg/m^2, not currently employed (73%) with an education level finished high school or higher (71%) (Table 8.1). Fourteen percent of these participants were on TJA waitlists at a

Figure 8.1 Flow diagram of study participants

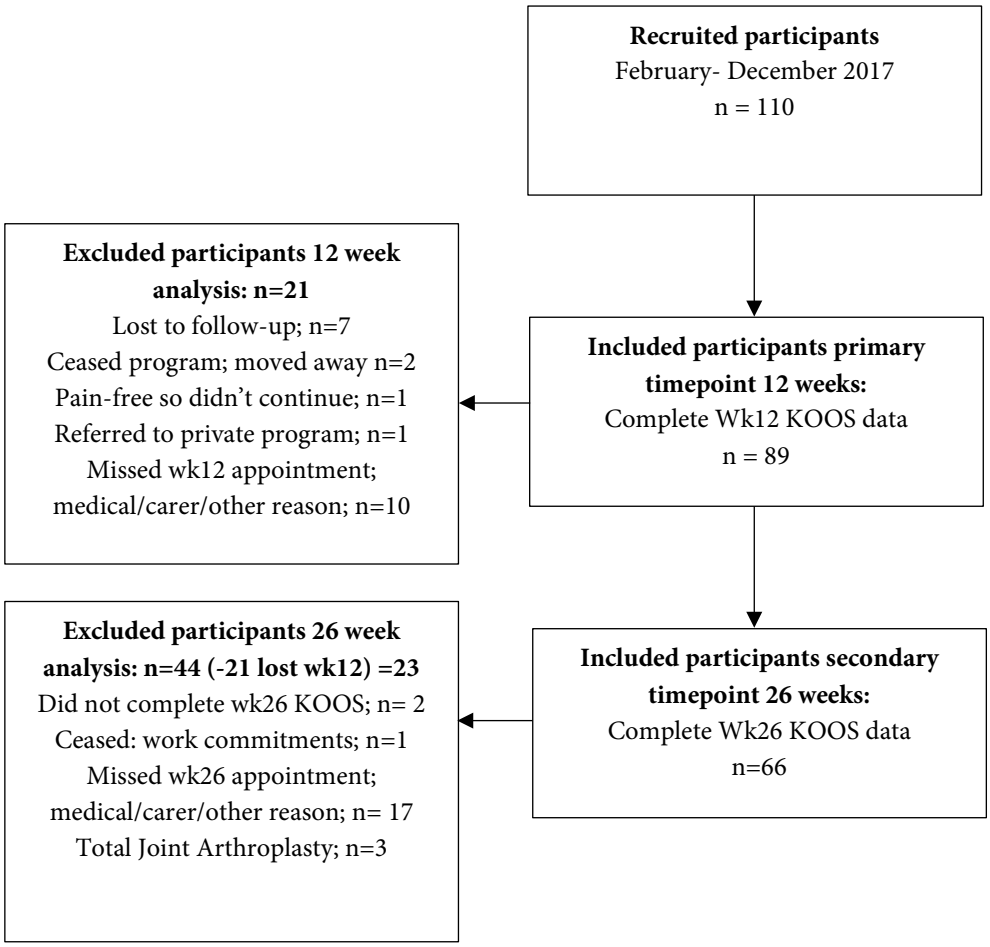


Table 8.1 Characteristics of participants.

Numbers are n (%) unless otherwise stated.

Characteristic	**n = 110**
Female	83 (75)
Age (years, mean [SD])	67.1 (10.0)
Residence‡: Lives alone	24 (27)
Currently employed^	30 (27)
Finished secondary school or higher∞	78 (71)
Did not finish secondary school°	32 (29)
On elective joint arthroplasty waitlist	15 (14)
BMIϱ, (mean [SD])	31.5 (6.2)
Pain NRS, (mean [SD], range 0-10)	3.8 (2.5)
TUG (units, mean [SD])	10.3 (4.4)
DASS Depression (mean [SD])	8.7 (9.9)
DASS depression ≥14	28 (25)
Anxiety (mean [SD])	6.6 (8.8)
Stress (mean [SD])	9.6 (9.6)
Number of comorbidities Low (0-1)	35 (32)
Number of comorbidities Moderate (2-3)	40 (36)
Number of comorbidities High (≥4)	25 (23)
Number of comorbidities Missing	10 (9)
Baseline PAM-13 Score (mean [SD])	59.1 (13.0)

PAM-13 Activation Level 1	19 (17)
PAM-13 Activation Level 2	20 (18)
PAM-13 Activation Level 3	56 (51)
PAM-13 Activation Level 4	15 (14)
KOOS converted to WOMAC Pain mean (mean [SD])	45.0 (19.2)
KOOS converted to WOMAC Function (mean [SD])	46.4 (18.1)

‡ Lives alone reported by participants. Living with others included living with able/non-able bodied person, hostel or aged care residential facility. ^ Currently employed includes participants who reported engaging in full/part-time paid work. ∞Included participants who reported finishing secondary school (final year), or university degree. ° Includes participants who did not finish secondary school and those who reported no formal schooling. KOOS: Knee injury and Osteoarthritis Outcome Score; WOMAC: Western Ontario and McMaster Universities Arthritis Index: 0 indicates no problems and 100 indicates extreme problems. Pain NRS: Numerical rating scale (0-10); BMIº: body mass index(kg/m^2). TUG: Timed Up and Go (seconds); DASS: Depression Anxiety Stress Scale: each subscale 0 best, 42 worst. PAM-13: Patient Activation Measure score 0 worst- 100 best.

public hospital. Baseline mean WOMAC pain and function scores were 45.0 (SD 19.2) and 46.4 (SD 18.1) respectively, and mean PAM-13 was 59.1 (SD 13.0). Of the 110 participants, 21 did not complete their 12-week reassessment (see Figure 8.1 for reasons for loss to follow up). The mean change in WOMAC pain and function scores at 12 weeks was -5.8 (95% CI 9.11, p<0.01) and -6.4 (95%CI -9.93, -3.39 p<0.01) points respectively (Table 8.2). Sixty-six participants returned for their 26-week assessment, (see Figure 8.1 reasons for loss to follow up). The mean change from baseline to 26-weeks WOMAC pain and function scores were -4.4 (95% CI -8.52, -0.27), p=0.04) and -4.4 (95%CI -8.05, -0.82, p=0.02) respectively. The mean improvement in baseline to 26-week PAM-13 scores was 5.8 (95%CI 1.89, 9.78, p<0.01) (Table 8.2).

Table 8.2 Change in PAM-13, pain and function baseline to 12 and 26- weeks

Variable	Baseline mean (SD) n=89	12-week mean (SD)	Difference 12-week - baseline (95% CI), P[v]	Baseline mean (SD) n=66	26-week mean (SD)	Difference 26-week - baseline (95% CI), P[v],
KOOS WOMAC pain	44.6 (19.5), n=89	38.8 (2.5)	**-5.8 (9.11, 2.57) p<0.01**	45.5 (20.2)	41.1 (22.5)	**-4.4 (-8.52, -0.27), p=0.04**
KOOS WOMAC function	45.5 (20.2) n=89	40.6 (20.8)	**-6.4 (-9.23, -3.49,) p<0.01**	46.2 (19.0)	41.8 (21.7)	**-4.4 (-8.05, -0.82,) p=0.02**
PAM-13	-	-	-	58.7 (13.3), n=66	64.5 (16.0)	**5.8 (1.89, 9.78) p<0.01**

Results in bold significant to alpha level 0.05. KOOS: Knee injury and Osteoarthritis Outcome Score; WOMAC: Western Ontario and McMaster Universities Arthritis Index: 0 indicates no problems and 100 indicates extreme problems. PAM-13: Patient Activation Measure score 0 worst- 100 best.

Unadjusted and baseline pain/function adjusted linear regression models

Tables 8.3 and 8.4 summarise the unadjusted and baseline pain/function adjusted linear regression models for change in pain and function at 12-weeks. PAM-13 scores were not significantly associated with changes in WOMAC pain and function at 12-weeks. Adjusted and unadjusted employment status and TUG were associated with WOMAC pain and function at 12-weeks. There was an association between age and change in WOMAC pain at 12-weeks which was not significant when adjusted for baseline.

The unadjusted and adjusted models for change in week-26 pain and function are presented in Tables 8.5 and 8.6. At the 26-week assessment, baseline PAM-13 were not associated with a change in WOMAC pain and function scores. Adjusted and unadjusted employment status was associated with a change in WOMAC function at 26-weeks. It was associated with a change in WOMAC pain only when adjusted for baseline. Baseline TUG results were associated with a change in WOMAC Function at 26-weeks when adjusted for baseline function.

Multivariable linear regression models

Baseline PAM-13 scores were not predictive of changes in WOMAC pain or function in the 12- or 26-week multivariable models (Tables 8.3, 8.4, 8.5, 8.6). Employment status (β=8.90 (95% CI 2.48, 15.30), p<0.01) and TUG (-1.32 (-1.20, -0.65), p<0.01) were predictive of week-12 WOMAC pain change scores when adjusted for baseline pain, PAM-13 and TJA waitlist status (F statistic= 9.065, p≤0.001, adjusted R^2= 0.32). Employment status (β= 11.01 (95%CI

5.05, 16.96), p<0.01) and TUG (β= -1.07 (95%CI -1.72, -0.42), p< 0.01) were also predictive of change in WOMAC function scores at 12-weeks when adjusted for baseline, PAM-13 scores, the DASS depression, age and TJA waitlist status (F statistic=7.063, p<0.01, adjusted R^2= 0.33). There were no significant predictors (p ≤0.05) in the multivariable models for change in baseline to 26-weeks WOMAC pain or function. There were no statistically significant correlations between baseline to 26-weeks change scores of PAM-13 and WOMAC pain (r= -0.14, p= 0.311) and function (r= -0.06, p= 0.636).

Table 8.3 Linear regression analyses; change in WOMAC pain Wk0 to Wk12

(n=89)

Variable	*β* (95% CI) (crude)	P	*β* (95% CI) (adjusted)*	P	Final multivariable model	P
Baseline PAM-13 score	-0.01 (-0.26, 0.24)	0.94	0.08 (-0.16, 0.33)	0.50	0.10 (-0.12, 0.31)	0.50
male	reference					
female	2.90 (-4.59, 10.39)	0.44	2.42 (-4.70, 9.55)	0.50	-	
Age (per year)	**-0.4 (-0.72, -0.04)**	**0.03**	-0.3 (-0.62, 0.05)	0.09	-	
		R^2= 0.05				
Didn't finish high school	reference					
High School/higher	5.18 (-2.23, 12.59)	0.17	5.21 (-1.86, 12.28)	0.15	-	
No paid employment	reference					
Paid employment	**10.00 (2.80, 17.21)**	**0.01**	**11.57 (4.81, 18.32)**	**<0.01**	**8.90 (2.48, 15.30)**	**0.006**
		R^2= 0.08		**R^2= 0.20**		
DASS Depression score <14	reference					
DASS Depression score ≥14	-1.52 (-9.16, 6.13)	0.69	-7.75 (-15.56, 0.06)	0.05	-	
Comorbidity count: 0-1	reference		reference		-	
Comorbidity count: 2-3	-0.57 (-9.00, 7.86)	0.89	-1.22 (-9.18, 6.73)	0.76	-	
Comorbidity count: ≥4	-4.55 (-14.33, 5.23)	0.36	-5.80 (-15.06, 3.46)	0.22	-	
Not on waitlist	reference		reference			
On waitlist	-8.27 (-18.54, 2.00)	0.11	-8.84 (-18.56, 0.88)	0.07	-5.07 (-14.02, 3.88)	0.26
TUG	**-0.97 (-1.69, -0.26)**	**0.01**	**-1.51 (-2.18, -0.84)**	**<0.01**	**-1.32 (-1.20, -0.65)**	**<0.01**
		R^2= 0.08		**R^2= 0.28**		
KOOS WOMAC pain	**0.26 (0.10, 0.42)**	**<0.01**			**0.40 (0.25, 0.55)**	**<0.01**

Results in **bold** significant to alpha level 0.05. KOOS: Knee injury and Osteoarthritis Outcome Score; WOMAC: Western Ontario and McMaster Universities Arthritis Index: 0 indicates no problems and 100 indicates extreme problems. PAM-13: Patient Activation Measure score 0 worst- 100 best. TJA: total joint arthroplasty; TUG: Timed Up and Go (seconds); DASS: Depression Anxiety Stress Scale: each subscale 0 best, 42 worst. P= p-value. *Adjusted for Wk0 WOMAC Pain score

Table 8.4 Linear regression analyses; change in WOMAC function Wk0 to Wk12

(n=89)

Variable	*β* (95% CI) (crude)	P	*β* (95% CI) (adjusted)*	P	Final multivariable model	P
Baseline PAM-13 score	0.03 (0.19, 0.25)	0.79	0.07 (-0.16, 0.29)	0.54	0.08 (-0.11, 0.28)	0.40
Male	reference					
Female	3.75 (-2.80, 10.30)	0.26	3.34 (-3.25, 9.92)	0.32	-	
Age	**-0.40 (-0.70, -0.11)**	**<0.01**	**-0.38 (-0.68, -0.08)**	**0.01**	-0.14 (-0.42, 0.14)	0.32
		R^2= 0.08		**R^2= 0.06**		
Didn't finish high school	reference		reference			
High School/higher	1.9 (-4.82, 8.39)	0.54	2.27 (-4.35, 8.89)	0.50	-	
No paid employment	reference		reference			
Paid employment	**14.44 (8.60, 20.28)**	**<0.01**	**14.72 (8.94, 20.51)**	**<0.01**	**11.01 (5.05, 16.96)**	**<0.01**
		R^2= 0.22		**R^2= 0.23**		
DASS Depression score <14	reference		reference			
DASS Depression score ≥14	-3.48 (-10.18, 3.23)	0.31	-6.71 (-14.11, 0.70)	0.08	-4.86 (-11.24, 1.52)	0.13
Low (0-1)	reference		reference			
Moderate (2-3)	1.10 (-5.99, 8.17)	0.76	0.93 (-6.12, 7.98)	0.79	-	
High (≥4)	-0.56 (-8.78, 7.66)	0.89	-1.41 (-9.69, 6.86)	0.74		
Not on waitlist	reference		reference			
On waitlist	-8.00 (-16.98, 1.00)	0.08	-8.25 (-17.21, 0.70)	0.07	-4.42 (-12.28, 3.45)	0.27
TUG	**-0.91 (-1.53, -0.28)**	**<0.01**	**-1.44 (-2.12, -0.76)**	**<0.01**	**-1.07 (-1.72, -0.42)**	**<0.01**
		R^2=0.09		**R^2= 0.17**		
KOOS WOMAC function	0.10 (-0.06, 0.27)	0.21			**0.31 (0.13, 0.48)**	**<0.01**

Results in **bold** significant to alpha level 0.05. KOOS: Knee injury and Osteoarthritis Outcome Score; WOMAC: Western Ontario and McMaster Universities Arthritis Index: 0 indicates no problems and 100 indicates extreme problems. PAM-13: Patient Activation Measure score 0 worst- 100 best. TJA: total joint arthroplasty; TUG: Timed Up and Go (seconds); DASS: Depression Anxiety Stress Scale: each subscale 0 best, 42 worst. P= p-value. *Adjusted for Wk0 WOMAC function score

Table 8.5 Linear regression analyses; change in WOMAC pain Wk0 to Wk26

(n=66)

Variable	**β (95% CI) (crude), P^{v}**	**β (95% CI) (adjusted)*, P^{v}**	**Final multivariable model, P^{v}**
Baseline PAM-13	0.03 (-0.28, 0.34), 0.85	0.09 (-0.22, 0.40), 0.60	0.10 (-0.20, 0.40), 0.49
male	reference	reference	
female	1.50 (-8.67, 11.67) 0.77	0.14 (-9.81, 10.10),0.98	-
Age	0.08 (-0.35, 0.51) 0.71	0.16 (-0.26, 0.57), 0.46	-
Not finished high school	reference	reference	
High School/higher	3.54 (-5.76, 12.83),0.45	3.54 (-5.45, 12.52),0.43	-
No paid employment	reference	reference	
Paid employment	7.69 (-2.06, 17.43),0.12	**10.14 (0.66, 19.62), 0.04**	9.10 (-0.36, 18.62), 0.06
DASS Depression <14	reference	reference	
DASS Depression ≥14	4.27 (-5.01, 13.55),0.36	0.53 (-9.26, 10.31), 0.92	-
Comorbidity count: 0-1	reference	reference	
Comorbidity count: 2-3	2.67 (-7.28, 12.62),0.59	3.45 (-6.16, 13.06), 0.48	-
Comorbidity count: ≥4	0.55 (-10.48,11.58), 0.92	0.2 (-10.46,10.80), 0.97	-
Not on TJA waitlist	reference	reference	
On TJA waitlist	3.08 (-10.42,16.57), 0.65	0.13 (-13.25, 3,52), 0.98	-
TUG	-0.41 (-1.48, 0.66), 0.44	-0.98 (-2.07, 0.12), 0.08	-0.94 (-2.03, 0.15), 0.09
KOOS WOMAC	**0.23 (0.03, 0.43), 0.03**		**0.34 (0.13, 0.55), 0.01**

Results in bold significant to alpha level 0.05. KOOS: Knee injury and Osteoarthritis Outcome Score; WOMAC: Western Ontario and McMaster Universities Arthritis Index: 0 indicates no problems and 100 indicates extreme problems. PAM-13: Patient Activation Measure score 0 worst- 100 best. TJA: total joint arthroplasty; TUG: Timed Up and Go (seconds); DASS: Depression Anxiety Stress Scale: each subscale 0 best, 42 worst. P^{v}= p-value.

*Adjusted for Wk0 WOMAC pain/function score

Table 8.6 Linear regression analyses; change in WOMAC function Wk0 to Wk26

(n=66)

Variable	**β (95% CI) (crude), P[v]**	**β (95% CI) (adjusted)*, P[v]**	**Final multivariable model, P[v]**
Baseline PAM-13	-0.01 (-0.28, 0.27), 0.97	0.03 (-0.24, 0.31), 0.82	0.06 (-0.21, 0.32), 0.67
male	reference	reference	
female	4.03 (-4.83, 12.88),0.37	3.35 (-5.47, 12.18), 0.45	-
Age	-0.04 (-0.41, 0.34), 0.84	0.01 (-0.38, 0.37), 0.98	-
Not finished high school	reference	reference	
High School/higher	4.15 (-4.06, 12.35), 0.32	5.08 (-3.03,13.20), 0.215	-
No paid employment	reference	reference	
Paid employment	**9.74 (1.40, 18.09), 0.02**	**10.25 (2.03, 18.48), 0.02**	8.10 (-0.24, 16.43), 0.06
DASS Depression <14	reference	reference	
DASS Depression ≥14	-4.30 (-12.41, 3.81), 0.29	-7.61 (-16.10, 0.87), 0.08	-6.20 (-14.47, 2.07), 0.14
Comorbidity count: 0-1	reference	reference	
Comorbidity count: 2-3	3.68 (-5.03, 12.39), 0.40	4.51 (-4.12, 13.14), 0.30	-
Comorbidity count: ≥4	1.50 (-8.16, 11.14), 0.76	0.80 (-8.73, 10.33), 0.87	
Not on TJA waitlist	reference	reference	
On TJA waitlist	-7.08 (-18.77, 4.62), 0.23	-8.90 (-20.57, 2.77), 0.13	-
TUG	-0.47 (-1.40, 0.46), 0.32	**-1.0 (-2.07, -0.02), 0.05**	-0.93 (-1.94, 0.09), 0.07
KOOS WOMAC	0.1 (-0.04, 0.34), 0.12		**0.31 (0.09, 0.53), 0.01**

Results in bold significant to alpha level 0.05. KOOS: Knee injury and Osteoarthritis Outcome Score; WOMAC: Western Ontario and McMaster Universities Arthritis Index: 0 indicates no problems and 100 indicates extreme problems. PAM-13: Patient Activation Measure score 0 worst- 100 best. TJA: total joint arthroplasty; TUG: Timed Up and Go (seconds); DASS: Depression Anxiety Stress Scale: each subscale 0 best, 42 worst. Pv= p-value.

*Adjusted for Wk0 WOMAC pain/function score

8.5. Discussion

The main purpose of this study was to examine the relationships between baseline patient activation and changes in pain and function following 12 weeks of an OACCP. Baseline PAM-13 scores were not associated with WOMAC pain and function in any of the models. Other variables such as employment status, TUG and the confounders depression, TJA waitlist status and age were independently associated with WOMAC pain and function scores at 12 weeks. Employment status and TUG results were associated with a change in WOMAC pain and function scores at 26 weeks. There were no significant predictors of change in symptoms at 26 weeks identified in the multivariable models. The mean change in PAM-13 scores was 5.8 (95%CI 1.89, 9.78) points following 26 weeks of the OACCP. The correlations between changes in PAM-13 and WOMAC pain/function scores at 26 weeks were weak.

Several previous studies have demonstrated that higher activation is associated with better longitudinal health outcomes in people living with chronic diseases (199-201). Most of the outcomes in these studies were comprised of general health behaviours such as healthy eating, managing medications or testing glucose levels. Few studies have examined the relationship between PAM-13 and aggregated change scores of outcomes to intervention.

One study examined the association between baseline PAM-13 scores and weight-loss following an incentive-based weight-loss intervention for obese individuals (225). Consistent with our findings, there was no association between PAM-13 and change scores (225). Conversely, another study reported that higher pre-operative PAM-13 scores predicted

greater improvement in Pain and Symptoms subscales of KOOS and Hip Disability and Osteoarthritis Outcome Scores (HOOS) following TJA for patients with advanced arthritis (226). These conflicting findings may be explained by the magnitude of the change scores. Changes in symptoms following non-surgical management are expected to be smaller than TJA particularly in people with end-stage OA (227). The change scores in our study may have been too subtle to detect an association.

Other independent predictors of changes in pain and function were identified. Employment status independently accounted for over 20% of the variability in change in 12-week WOMAC Function scores and 20% of the variability of change in pain scores when adjusted for baseline. Employment status has been associated with age and the severity of OA symptoms in previous studies (228, 229). It could be argued that employment status was acting as a proxy for disease severity and/or age in this analysis, however, we adjusted the models for these variables, so this should not be the case. The relationship between employment status and improved outcomes following OAMPs warrants further attention.

Timed Up and Go accounted for 28% and 17% of the variability in change of WOMAC pain and function scores at 12 weeks respectively, when adjusted for baseline. A possible explanation for this finding is that people with higher functional ability may have found it easier to fully engage with their exercise program, and hence were more able to achieve the potential beneficial treatment effects associated with the intervention. Better TUG times may reflect higher levels of physical activity and/or exercise of participants in their everyday lives.

Previous experience with exercising has been identified in a scoping review as a facilitator of exercise participation (221); this may also have improved the engagement of these higher functioning participants with the program and led to better treatment effects.

The mean change in PAM-13 scores following 26 weeks of the OACCP (5.8 (95%CI 1.89, 9.78) points) was small considering that the PAM-13 is a 100- point scale. This change was similar to the mean change reported in a previous randomised trial (200). This trial included participants with at least one of several chronic diseases: arthritis; diabetes; hypertension; heart disease; chronic obstructive pulmonary disease, or hyperlipidaemia. Participants randomised to the Chronic Disease Self-Management Program reported a mean change of 4.7 PAM-13 points at 26-weeks (200). However, the authors identified two trajectories of change in activation: i) an increased PAM trajectory, and ii) a stable PAM trajectory. The participants with increased PAM trajectory membership reported a mean change in PAM-13 of 15.4 points at six months, while those with the stable PAM trajectory had a slightly worse mean PAM-13 scores at six months (compared to baseline). Although all study participants increased their desirable self-management behaviours, the participants with increased PAM trajectory membership demonstrated greater improvements across more self-management behaviours (200). This present study did not examine different trajectories in change of PAM-13 scores following the OACCP. This could be explored further in larger cohorts of OACCP participants.

Changes in PAM-13 scores were not correlated with changes in pain and function. It is possible that the symptomatic benefits from lifestyle interventions requiring substantial behaviour change take longer to result in improvements that are noticeable and important to patients. The analyses at 12 and 26 weeks may not have allowed adequate time for participants to harness their activation to change behaviour and experience noticeable changes in pain and function. In addition, to reduce participant burden attributed to multiple tests, PAM-13 was not routinely administered at 12-weeks. We were unable to determine whether changes in PAM-13 scores at 12 weeks were correlated with changes in pain and function. Also, there was loss to follow up throughout this clinical cohort study (follow-up rates were 81% and 60% at 12- and 26-weeks respectively). This may have contributed to the lack of association between baseline measures and changes in symptoms at the 26-week time point.

Although improvements in pain and function scores were achieved following participation in OACCPs, these were not associated with PAM-13 baseline or change scores. Other independent predictors of changes in symptoms including employment status and TUG were more helpful predictors of change in pain and function.

Acknowledgement

We would like to acknowledge the participants of the OA chronic disease management programs at Royal North Shore and Mount Wilga Hospitals who graciously completed the PAM-13 along with the other patient-reported outcomes used as part of the program.

CHAPTER NINE

Chapter Nine: Thesis discussion and future directions

9.1. Summary of Thesis findings

This aim of this Thesis was to identify clinical characteristics that were associated with positive responses and worsening following participation in the OACCP. A range of disease-specific, psychological, physical and demographic variables were examined. However, an overall finding was that the variables that were significantly associated with OACCP outcomes only explained a small amount of the variability of the outcome.

This Thesis has confirmed previous evidence that not all participants respond favourably following an OAMP (27, 29, 30). The mean improvements of people with knee OA, who had participated in 12-weeks of the OACCP in Chapter Eight, was approximately 6/100 points for both WOMAC Pain and Function scores respectively. Although these improvements were small, they were comparable to the results of other OAMPs (30, 34). In Chapter Four only 28% Of OACCP participants reached the MCID for 'response' (39) . While this response rate seems low, it aligns with the results reported in most of the previous literature (29, 34). The exception was the ESCAPE-Knee Pain program where 61% and 54% of participants reached the threshold for response at six weeks and six months respectively. An important difference in the ESCAPE-Pain study was that the threshold of change in WOMAC scores used to indicate response was lower than the one used in Chapter Four: this probably explains the higher proportion of responders compared with our study.

The low response rates found in this Thesis and in the literature according to disease-specific measures such as the WOMAC, prompts consideration of whether these variables are the most meaningful outcomes to measure the effectiveness of OAMPs. It was argued in Chapter Six (41) that the measurement of OA self-management capabilities is important to indicate the success of OAMPs. The authors of a systematic review on the effectiveness of OA self-management education programs reported a "mismatch between the aims of self-management education programs and the outcomes used to measure success" (15). They argued further that pain is often used as the main outcome for non-surgical treatments for OA, and although a reduction in pain is desirable for patients, it may not fully reflect the aims of treatments centred on enabling patients to self-manage their condition (15). Most often, the main aim of OAMPs is to support patients to live well with their OA, despite their ongoing pain and other symptoms. Hence, it is arguable that self-management capability is an important indicator of efficacy (15).

Osteoarthritis self-management knowledge, attitudes and capabilities have not been widely assessed in previous OAMP literature. Following the systematic review in Chapter Six, we were unable to recommend a suitable instrument with demonstrated adequate measurement properties to measure these constructs. In the absence of a suitable instrument, the measurement properties of the PAM-13 were evaluated in a cohort of OACCP participants in Chapter Seven. The PAM-13 was then evaluated further in a longitudinal cohort in Chapter Eight.

The study in Chapter Eight demonstrated that patient activation is positively modifiable in an OA population, potentially more so than for several other chronic conditions (200). This is an important finding because the improvements in PAM-13 scores have previously been associated with increased desirable self-management behaviours (200). We did not investigate the relationship between patient activation and desirable health-related behaviours in this Thesis. However, an important future research question may be to determine whether longitudinal PAM-13 scores and specific health behaviours, such as meeting the requirements of physical activity recommendations or adhering to prescribed daily calorific intake to achieve weight loss, concurrently improve in people participating in OAMPs. If PAM-13 scores were found to reflect changes in health-specific behaviours, this would provide important evidence to support the construct validity of PAM-13 in this population.

We hypothesised that people with higher patient activation at enrolment in the OACCP would have better outcomes. In Chapter Eight neither patient activation nor change in activation, were associated with changes in WOMAC pain and function following participation in the OACCP. This finding was surprising, especially given the findings of a previous study on the Good Living with Osteoarthritis Denmark (GLA:D) program. The GLA:D study reported that self-efficacy measured on the Arthritis self-efficacy scale (ASES) at three months; and the change in ASES from baseline to three months; were significantly associated with improvements in pain at 12 months (130). These conflicting results may be related to differences in the specificity of the constructs measured by the PAM-13 and the ASES. The PAM-13 items are not targeted to specific OA management tasks, but ask about

the management of general health e.g. “I am confident I can help prevent or reduce problems associated with my health”. On the other hand, self-efficacy refers to the belief that one can achieve a specific task or behaviour (230). The items of the ASES start with the prefix: “How certain are you that you can…” and then list specific tasks to manage arthritis e.g. “make a large reduction in your arthritis pain by using methods other than taking extra medication?” (191). The ASES contains specific questions that ask about OA self-management. These items may have a stronger statistical relationship with improvements in measures of pain and function because they relate more specifically to OA. Whereas, the PAM-13 items are directed at management of general health (42). This may explain why baseline and changes in PAM-13 scores were not associated with changes in pain and function in Chapter Eight.

Several other characteristics were significantly associated with changes in pain and function following the OACCP. Some of these variables were non-modifiable such as gender and index joint (hip vs knee), and other variables were potentially modifiable such as TJA waitlist status, functional performance (TUG) and employment status.

Non-modifiable characteristics

i) Index joint

Participants who presented with the knee as the index joint had twice the log odds of being responders compared to those with hip OA in Chapter Four (39). A possible explanation for this result was that the size of the treatment effects observed following exercise were higher for people with knee OA compared to those with hip OA. Evidence from systemic reviews

suggests that the clinical efficacy of exercise differs for people with hip OA compared with knee OA for both pain and physical function. In a meta-analysis investigating interventions for knee OA, the pooled treatment effect was 0.5 standardised mean difference immediately following treatments (16). For those with hip OA, this effect was smaller (0.38) and the confidence intervals was less precise, indicating more variability in the treatment effect (17). Based on these findings, it is not surprising that people with hip OA had lower odds of being considered responders in Chapter Four (39).

There are potential clinical implications of participants with knee OA having twice the log odds of being responders following 26-weeks of the OACCP. First, this finding needs to be replicated in other large sample size clinical cohort studies. If this finding is confirmed in further research, this evidence may alter the expectations of participants and their clinicians regarding the magnitude of symptomatic relief that participants with hip OA achieve. This is not to say that people with hip OA will worsen. There was no significant association between index joint and worsening in Chapter Five (40). Furthermore, several studies have demonstrated that index joint was not a stable predictor of response (29, 30, 34). This underscores the need for caution when interpreting differences between knee and hip OA responders. Given the evidence for significant, small to moderate effects on pain and function derived from exercise therapy for hip OA (17), it would be inappropriate to exclude people with hip OA from entry to OACCP based on their index joint.

ii) Sex

Based on findings from Chapter Four, female participants had twice the log odds of being responders compared to males (39). Sex was not significantly associated with worsening in Chapter Five (40), nor with a change in WOMAC pain and function in Chapter Eight of this Thesis. The evidence in the literature of the association between sex and outcomes of OAMPs is conflicting (29, 30, 34). The clinical utility of using sex as a predictor of response is limited. It would be considered unacceptable to triage participants for entry to the OACCP based on sex even if there were stronger evidence to support this finding.

Modifiable factors

i) Total joint arthroplasty waitlist status

Total joint arthroplasty (TJA) waitlist status was significantly associated with increased odds of worsening in Chapter Five following 26 weeks participation in the OACCP (OR 1.91 [95%CI 1.04, 3.51]) when adjusted for baseline 6-minute-walk-test results (40). However, this model only explained 5% of the variance in demonstrating worsening (40), suggesting that there are other unknown factors that need to be considered. In addition, the relationship between TJA waitlist status and worsening changed depending on the definition of worsening that was used (40). There was a similar finding in a previous longitudinal observational cohort study of people on hip and knee joint replacement waitlists. Using similar thresholds of clinically important worsening and improvement in WOMAC scores as Chapters Four and Five, half of the cohort reported unchanged WOMAC score, approximately a 25% reported worsening and 25% reported improvement while waiting for TJA (156). In addition, participants were asked transition questions about self-perceived change in wellbeing

(including: pain, fatigue, overall quality of life, overall health). Over half of the participants reported they were 'worse' on the transition questions. An obvious difference between this cohort and the one in Chapter Five was that the participants did not receive any intervention. However, the study did identify different trajectories of people on TJA waitlists; and similar to Chapter Five, the proportion of participants who reported they were 'worse' was dependent on the threshold used to indicate worsening (156).

In Chapter Eight TJA waitlist status was not an independent predictor of change in WOMAC pain and function. However, it was included as a confounder in the final multivariable models for WOMAC pain and function at 12-weeks. When TJA waitlists status was removed from the model, the regression coefficients of the other variables changed (>10%) such that it suggested TJA waitlist status bore some influence on the relationships between the other variables in the model. TJA waitlist status was included in the final multivariable models of the studies in both Chapters Five and Eight, and both studies suggested that there is a relationship between TJA waitlist status and outcomes following the OACCP (40). This evidence lends weight to the recommendation in OA management guidelines that TJA should only be considered in cases where the patient has disabling symptoms, and all non-surgical treatments (non-pharmacologic and pharmacologic) have been exhausted (13). Further, early recruitment for participation in OAMPs, prior to referral for consultation with orthopaedic surgeons, may enable people with OA to derive the greater benefits from non-surgical treatments.

ii) Functional performance measures

The TUG assesses a multitask activity that includes transitions between sitting, standing and walking (142). Baseline TUG was associated with of changes in pain and function, explaining 28% and 17% of the variability in change of WOMAC pain and function scores at 12 weeks respectively, when adjusted for baseline. Although the TUG is a test of functional performance for hip and knee OA recommended by OARSI (142), there is little literature available detailing biomechanical forces at the knee during this test. One previous study from, over 30 years ago, demonstrated that standing from sitting causes loading across the different joint compartments of the knee (231). The forces at the patellofemoral joint were approximately 15x bodyweight which were almost twice the magnitude of those at the tibiofemoral joint (231). By vigorously loading the different compartments of the knee, the TUG may provide a comprehensive outcome that reflects the overall health of the OA knee. This may explain why the TUG was associated with changes in pain and function found in Chapter Eight.

The ability of the TUG to discriminate OA functional severity may have important clinical implications. Participants with higher functional ability, indicated by lower TUG times may find it easier to fully engage with their OACCP interventions, especially their exercise program, and hence be enabled to achieve the potential beneficial treatment effects associated with the interventions. Assessment using the TUG may therefore identify patients most likely to respond to OACCP interventions and enable OAMP resources to be targeted towards those individuals. While this hypothesis would need to be tested in future research, this

finding could modify patient and health professional expectations regarding the likely benefits derived from OACCP participation. Our findings suggest that better improvements in pain and function could be expected from participants with higher baseline functional performance measured using TUG.

iii) Employment status

Another potentially modifiable variable and independent predictor of change in WOMAC pain and function scores in Chapter Eight was employment status. Employment status independently accounted for over 20% of the variability in change in 12-week WOMAC Function scores and 20% of the variability of change in pain scores when adjusted for baseline in Chapter Eight. This relationship between employment status, pain and function has not been previously examined within the context of an OAMP. One possible explanation for this finding is that employment was acting as a proxy for OA symptom severity. A previous Australian population-based study suggested that people living with lower limb joint disease who had very low health quality of life indicated by their WOMAC scores, had more than three times the odds of not being engaged in paid employment (229). In our study, people who were engaged in paid employment may have presented to the OACCP with lower levels of OA disease severity; and being at an earlier stage of the condition they may have responded to treatment more favourably. The relationship between employment status, disease severity and improved outcomes following the OACCP warrants further attention in longitudinal cohort studies.

iii) Depressive symptoms

The presence/absence of depressive symptoms was not significantly associated with either improvement or worsening in Chapters Four (39), Five (40) or Eight. Our findings contrasted with some previous studies. The absence of depression was identified previously as a predictor of response or improvement following a 3-4 week inpatient multimodal rehabilitation intervention (34) and a weight loss program in overweight veterans with knee OA (80). A "depressive knee OA" phenotype was identified in previous two large cross-sectional cohort studies: people classified as the depressive knee OA phenotype reported significantly worse self-reported pain and function outcomes than most of the other phenotypes (232).

The absence of any significant association between baseline depressive symptoms and OACCP clinical outcomes in this Thesis may be explained by the effect of the OACCP intervention on depression itself. All participants of the OACCP were screened for symptoms of mental health conditions using the Depression, Anxiety and Stress Scale (DASS) (159). The participants of the OACCP who indicated that they had experienced depressive symptoms on the DASS depression subscale, were referred for consultation with a social worker (or less frequently a psychologist) as required.

The evidence for the clinical efficacy of treatment for OA sufferers with comorbid depression is conflicting. There is some RCT evidence that the treatment of depression in people with arthritis reduced their pain, depressive symptoms and improved function and quality of life

(92). Subsequent analyses found that depression management (compared to usual care) was effective in reducing pain severity among people with arthritis with lower levels of baseline pain, but these benefits were not observed in people with higher initial pain severity (233). A more recent cluster RCT in primary care practices in the United Kingdom compared two methods of screening for and managing depressive symptoms. One group of general practices were randomised to a point-of-care screening form to identify people presenting with OA who were experiencing comorbid symptoms of depression and/or anxiety; and provided management of depression/anxiety when required. The other group of practices were randomised to usual care. Compared with usual care, the group who received the screening and depression/anxiety intervention reported worse pain outcomes and no improvement in their depression/anxiety symptoms (234). Together, these studies demonstrate that the relationship between the management of depression and the impact on OA symptoms remains unclear. Further RCTs are required to determine the effects of treatment of depression on OA symptoms.

iv) Number of comorbidities

The presence of depression and other comorbidities have been previously associated with poorer health-related quality of life for OA patients waiting for TJA (171). Low comorbidity counts were associated with improvements in OA symptoms following a 3-4 week inpatient multimodal rehabilitation intervention for hip and knee OA (34). In contrast, our study found higher comorbidity counts were not significantly associated with improving or worsening following the OACCP in Chapters Four or Five of this Thesis (39, 40). The

OACCP model of care aims to include assessment, education and interventions to support the management of comorbidities (134). The assessment and management of comorbidities, within the context of this chronic care model, may have reduced the negative impact of higher comorbidity counts on OACCP outcomes. This is an important clinical finding as it suggests that OACCP may be considered suitable for anyone with any number of concomitant conditions.

9.2. Strengths and limitations of this Thesis

The clinical cohort designs of the studies of this Thesis brought both strengths and limitations to the findings of this Thesis. A clinical cohort study generally includes a sample of individuals who are actively receiving health care in a clinic setting, for a specified condition, and uses their health data to assess outcomes. The data collection intervals are determined by the nature of the healthcare services provided (235). The clinical cohorts in this Thesis were comprised of real-life OACCP clinics and the data were collected at program assessments according to standard clinical practice. Although this had implications on the timing and completeness of the data, this was a strength of our study as it also increased the applicability of the study findings to real-world clinical settings (235).

It is important, however, to consider the nature of the real-world settings that the clinical cohorts were drawn from. The study data were drawn from a limited number of OACCPs in socio-demographically similar regions of NSW, Australia. The resulting study samples represented homogenous, metropolitan geographic areas of Australia with higher socio-

economic and educational status. This was a limitation of the findings of this Thesis. A previous multicentre randomised controlled trial suggested that sociodemographic and geographic characteristics have an impact on the availability and uptake of some components of OA treatment (236). The homogenous samples used in this Thesis may have diminished the applicability of these studies and this needs consideration when interpreting the results.

A further limitation is the applicability of our findings to people living with hip OA due to the low number of hip OA participants in our studies. There were much larger proportions of people with knee OA, compared to hip OA in the studies reported in Chapters Four (39), Five (40), and Seven. Chapter Eight only included people with the knee as the index treatment joint. The lower proportions of participants with hip OA in Chapters Four, Five and Seven reflected the lower prevalence of hip compared with knee OA (237) and referral patterns of the OACCP. The low hip OA participation rate should be considered when interpreting the results of this Thesis in relation to hip OA.

Despite some limitations to the applicability of the findings of the studies, the clinical cohort study design provided an important strength to this Thesis. This design enabled recruitment of large sample sizes for the studies in Chapters Four, Five and Eight. This allowed for adequate statistical power to detect associations between the variables and the dependent outcomes, which imbues confidence in the results of the studies overall. Although the study design enabled large study sample sizes, it was also prone to problems with loss-to-follow-up that is typical of clinical practice (235). Participants left the OACCP for various reasons

including: they reached the top of the TJA waitlists and underwent surgery; withdrew on medical advice; moved interstate; or suffered illness or death. The earlier studies of this Thesis (Chapters Four (39) and Five (40)) saw larger losses to follow-up than the later studies. This could be attributed to larger proportions of participants referred to the OACCP from TJA waitlists at that earlier time.

This censoring due to TJA may have introduced selection bias because the participants with the worst disease severity may have been lost to follow up as they left the OACCP to have TJA. The participants who attended the OACCP at the later data collection timepoints may have been those who were responding most favourably to the intervention. This may have resulted in overestimation of the mean changes in symptoms following the OACCP for the population. However, we used techniques to mitigate the risk of selection bias such as imputation of data for participants lost to follow up in these chapters (39, 40).

A further strength of our studies was that the follow-up periods for the longitudinal studies were clearly defined and clinically meaningful. The follow-up period for the primary analyses in Chapters Four and Five was at 26-weeks. This was chosen to allow for sufficient opportunity for OACCP participants to achieve changes in their symptoms to achieve the MCID used to indicate response or worsening (32). The follow-up time chosen for the primary analysis in Chapter Eight was at 12-weeks. This timepoint was chosen considering that the changes in symptoms achieved in the early months of a program may greatly influence the long-term success of participants. The association between early success and

better long-term outcomes was demonstrated in a previous study of GLA:D participants (130). A further advantage of using the 12-week follow up was it avoided some of the loss to follow-up that was seen at the 26-week follow-up timepoint in Chapters Four and Five. In maximizing the sample of complete data by using the earlier 12-week timepoint, the models had more statistical power to examine the associations between baseline characteristics and outcomes of the OACCP because we had a larger sample size.

The variables examined in relation to their association with OACCP outcomes in Chapters Four, Five and Eight were identified *'a priori'* through literature and peer review. A broad mix of disease, psychological, physical and demographic variables were included, while giving due consideration to the number of variables examined, to avoid overfitting the regression models with excessive degrees of freedom (53, 55). This is considered an important strength: however, we acknowledge a potential limitation of the clinical cohort study design was that the effect of these variables on treatment outcomes were not compared with a control group. It was only possible to identify prognostic factors that were associated with a treatment outcome. To truly identify subgroups of people who experience a different effect of treatment compared with another group, it would be necessary to use a RCT trial design (238). An RCT was not a feasible study design for this Thesis and was unnecessary for the exploratory nature of the research question. We understand that the variables that explained larger amounts of the variation in outcomes (i.e. employment status and TUG), need to be validated in similar but different samples of OACCP participants and that ideally, this would be compared with people not participating in such programs (239).

A further potential methodological limitation of this Thesis was the categorisation of some of the independent variables. Categorising data inevitably leads to loss of information; however, it was necessary to reflect the clinical utility of some outcome measures by using established cut-offs taken from clinical practice. For example, DASS depression scores were dichotomised to indicate the presence/absence of depressive symptoms in Chapters Four (39), Five (40) and Eight. In treating DASS depression as a categorical variable, we may have diminished the ability of the models to detect relationships with the outcome. There were also limitations regarding the methods used to describe and quantify comorbidities in this Thesis. These were limited to simple counts of diagnosed conditions that did not take into consideration the seriousness or impact each comorbidity may have on overall health. This may have overlooked implications that certain comorbidities may have for people with OA e.g. cardiovascular disease, diabetes or obesity (11).

9.3. Future directions

Osteoarthritis is now widely recognised as an heterogenous, inflammatory disease with distinct observable characteristics, also known as phenotypes (240). A well-documented limitation of extant clinical trials investigating the effectiveness of OA treatments is that sample populations 'lump' patients with different pathoaetiologies together and expect to find a significant overall treatment effect (238, 241-243). Grouping together subgroups which both respond and don't respond to interventions in RCTs potentially masks the overall treatment effect and this may lead to diminished or even negative findings from the RCT (242). This

phenomenon probably contributed to the relatively underwhelming small to moderate treatment effects of interventions recommended as effective therapies for OA (68).

As our understanding of OA grows, it is crucial that this knowledge is used to conduct research that enable clinicians and researchers to better target treatments to individuals according to their clinical presentation (238).Doing so should enable us to direct limited clinical resources to individuals most likely to derive benefit from them. Whilst our studies investigated associations between patient reported outcomes, and subgroups identified using functional performance variables, personal factors, demographic and psychosocial variables, there have been a range of alternative clinical features studied in the literature that could potentially indicate different subgroups of OA. Suggested features include: the site (joint compartments), extent and structural subtypes of OA, injury history, range of motion, radiographic features (e.g. effusion/synovitis and bone marrow lesions), and specific comorbidities (e.g. cardiovascular disease) (238). There has also been great interest in phenotypes defined by cellular features and biomarkers such as: inflammatory and autoimmune markers; cell senescence; metabolic, genetic and hormonal features; and signs associated with mechanical overload (242, 243).

Work emerging in this field includes a recent systematic review that examined phenotypes of knee OA and their relationship with clinical and structural outcomes (244). The review identified substantial heterogeneity in the types of variables used to define OA phenotypes in observational studies (244). Once the characteristics of these subgroups have been identified

and validated, the associations of subgroup memberships with clinical outcomes following OACCP participation will need to be determined.

9.4. Final Remarks

- This Thesis identified some significant predictors of outcomes following the OACCP which included gender, index joint and TJA waitlist status. However, the models were not adequate to explain much of the variability in these outcomes.
- In the absence of strong predictors of improvement or worsening, we sought to examine the association between OA self-management attitudes and capabilities and outcomes.
- There was no instrument recommended to measure OA self-management attitudes and capabilities. We conducted a systematic review that revealed a paucity of measurement property evidence for instruments measuring this construct.
- We evaluated the measurement properties for PAM-13 using Rasch analysis in an OA population.
- Patient activation was not significantly associated with changes in measures of pain and function, however there were other characteristics that were associated with these outcomes.

References

1. Hawker GA, Stewart L, French MR, Cibere J, Jordan JM, March L, et al. Understanding the pain experience in hip and knee osteoarthritis--an OARSI/OMERACT initiative. Osteoarthritis Cartilage. 2008;16(4):415-22.
2. Allen KD, Golightly YM. State of the evidence. Curr Opin Rheumatol. 2015;27(3):276-83.
3. Hootman JM, Helmick CG, Barbour KE, Theis KA, Boring MA. Updated Projected Prevalence of Self-Reported Doctor-Diagnosed Arthritis and Arthritis-Attributable Activity Limitation Among US Adults, 2015-2040. Arthritis Rheumatol. 2016;68(7):1582-7.
4. Neogi T, Zhang Y. Epidemiology of osteoarthritis. Rheum Dis Clin North Am. 2013;39(1):1-19.
5. Felson DT, Naimark A, Anderson J, Kazis L, Castelli W, Meenan RF. The prevalence of knee osteoarthritis in the elderly. The Framingham Osteoarthritis Study. Arthritis Rheum. 1987;30(8):914-8.
6. Jordan JM, Helmick CG, Renner JB, Luta G, Dragomir AD, Woodard J, et al. Prevalence of knee symptoms and radiographic and symptomatic knee osteoarthritis in African Americans and Caucasians: the Johnston County Osteoarthritis Project. J Rheumatol. 2007;34(1):172-80.
7. Haugen IK, Englund M, Aliabadi P, Niu J, Clancy M, Kvien TK, et al. Prevalence, incidence and progression of hand osteoarthritis in the general population: the Framingham Osteoarthritis Study. Ann Rheum Dis. 2011;70(9):1581-6.
8. Cross M, Smith E, Hoy D, Nolte S, Ackerman I, Fransen M, et al. The global burden of hip and knee osteoarthritis: estimates from the global burden of disease 2010 study. Ann Rheum Dis. 2014;73(7):1323-30.
9. Xie F, Kovic B, Jin X, He X, Wang M, Silvestre C. Economic and Humanistic Burden of Osteoarthritis: A Systematic Review of Large Sample Studies. Pharmacoeconomics. 2016;34(11):1087-100.
10. Duffield SJ, Ellis BM, Goodson N, Walker-Bone K, Conaghan PG, Margham T, et al. The contribution of musculoskeletal disorders in multimorbidity: Implications for practice and policy. Best Pract Res Clin Rheumatol. 2017;31(2):129-44.
11. Parkinson L, Waters DL, Franck L. Systematic review of the impact of osteoarthritis on health outcomes for comorbid disease in older people. Osteoarthritis Cartilage. 2017;25(11):1751-70.
12. Nelson AE, Allen KD, Golightly YM, Goode AP, Jordan JM. A systematic review of recommendations and guidelines for the management of osteoarthritis: The chronic osteoarthritis management initiative of the U.S. bone and joint initiative. Semin Arthritis Rheum. 2014;43(6):701-12.

13. Meneses SR, Goode AP, Nelson AE, Lin J, Jordan JM, Allen KD, et al. Clinical algorithms to aid osteoarthritis guideline dissemination. Osteoarthritis Cartilage 2016;24(9):1487-99.
14. Du S, Yuan C, Xiao X, Chu J, Qiu Y, Qian H. Self-management programs for chronic musculoskeletal pain conditions: a systematic review and meta-analysis. Patient Educ Couns. 2011;85(3):e299-310.
15. Kroon FP, van der Burg LR, Buchbinder R, Osborne RH, Johnston RV, Pitt V. Self-management education programmes for osteoarthritis. The Cochrane database of systematic reviews. 2014;1(1):CD008963.
16. Fransen M, McConnell S, Harmer AR, Van der Esch M, Simic M, Bennell KL. Exercise for osteoarthritis of the knee: a Cochrane systematic review. Br J Sports Med. 2015;49(24):1554-7.
17. Fransen M, McConnell S, Hernandez-Molina G, Reichenbach S. Exercise for osteoarthritis of the hip. Cochrane Database Syst Rev. 2014(4):CD007912.
18. Christensen R, Bartels EM, Astrup A, Bliddal H. Effect of weight reduction in obese patients diagnosed with knee osteoarthritis: a systematic review and meta-analysis. Ann Rheum Dis. 2007;66(4):433-9.
19. Zhang L, Fu T, Zhang Q, Yin R, Zhu L, He Y, et al. Effects of psychological interventions for patients with osteoarthritis: a systematic review and meta-analysis. Psychol Health Med. 2018;23(1):1-17.
20. Jones A, Silva PG, Silva AC, Colucci M, Tuffanin A, Jardim JR, et al. Impact of cane use on pain, function, general health and energy expenditure during gait in patients with knee osteoarthritis: a randomised controlled trial. Ann Rheum Dis. 2012;71(2):172-9.
21. Raja K, Dewan N. Efficacy of knee braces and foot orthoses in conservative management of knee osteoarthritis: a systematic review. Am J Phys Med Rehabil. 2011;90(3):247-62.
22. Allen KD, Choong PF, Davis AM, Dowsey MM, Dziedzic KS, Emery C, et al. Osteoarthritis: Models for appropriate care across the disease continuum. Best Pract Res Clin Rheumatol. 2016;30(3):503-35.
23. Runciman WB, Hunt TD, Hannaford NA, Hibbert PD, Westbrook JI, Coiera EW, et al. CareTrack: assessing the appropriateness of health care delivery in Australia. Med J Aust. 2012;197(2):100-5.
24. Basedow M, Esterman A. Assessing appropriateness of osteoarthritis care using quality indicators: a systematic review. J Eval Clin Pract. 2015;21(5):782-9.
25. Basedow M, Williams H, Shanahan EM, Runciman WB, Esterman A. Australian GP management of osteoarthritis following the release of the RACGP guideline for the non-surgical management of hip and knee osteoarthritis. BMC Res Notes. 2015;8:536.
26. Paskins Z, Sanders T, Croft PR, Hassell AB. The Identity Crisis of Osteoarthritis in General Practice: A Qualitative Study Using Video-Stimulated Recall. Ann Fam Med. 2015;13(6):537-44.
27. Hurley MV, Walsh NE, Mitchell H, Nicholas J, Patel A. Long-term outcomes and costs of an integrated rehabilitation program for chronic knee pain: a pragmatic, cluster randomized, controlled trial. Arthritis Care Res (Hoboken). 2012;64(2):238-47.

28. Skou ST, Roos EM. Good Life with osteoArthritis in Denmark (GLA:D): evidence-based education and supervised neuromuscular exercise delivered by certified physiotherapists nationwide. BMC Musculoskelet Disord. 2017;18(1):72.
29. Snijders GF, van den Ende CH, van den Bemt BJ, van Riel PL, van den Hoogen FH, den Broeder AA, et al. Treatment outcomes of a Numeric Rating Scale (NRS)-guided pharmacological pain management strategy in symptomatic knee and hip osteoarthritis in daily clinical practice. Clin Exp Rheumatol. 2012;30(2):164-70.
30. Snijders GF, den Broeder AA, van Riel PL, Straten VH, de Man FH, van den Hoogen FH, et al. Evidence-based tailored conservative treatment of knee and hip osteoarthritis: between knowing and doing. Scand J Rheumatol. 2011;40(3):225-31.
31. Pham T, van der Heijde D, Altman RD, Anderson JJ, Bellamy N, Hochberg M, et al. OMERACT-OARSI initiative: Osteoarthritis Research Society International set of responder criteria for osteoarthritis clinical trials revisited. Osteoarthritis Cartilage. 2004;12(5):389-99.
32. Angst F, Aeschlimann A, Michel BA, Stucki G. Minimal clinically important rehabilitation effects in patients with osteoarthritis of the lower extremities. J Rheumatol. 2002;29(1):131-8.
33. Lamb SE, Toye F, Barker KL. Chronic disease management programme in people with severe knee osteoarthritis: efficacy and moderators of response. Clin Rehabil. 2008;22(2):169-78.
34. Weigl M, Angst F, Aeschlimann A, Lehmann S, Stucki G. Predictors for response to rehabilitation in patients with hip or knee osteoarthritis: a comparison of logistic regression models with three different definitions of responder. Osteoarthritis Cartilage. 2006;14(7):641-51.
35. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. Patient Educ Couns. 2002;48(2):177-87.
36. Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for Outcome Measures in Rheumatology. J Rheumatol. 1998;25(2):198-9.
37. Tugwell P, Boers M, D'Agostino MA, Beaton D, Boonen A, Bingham CO, 3rd, et al. Updating the OMERACT filter: implications of filter 2.0 to select outcome instruments through assessment of "truth": content, face, and construct validity. J Rheumatol. 2014;41(5):1000-4.
38. Eyles J, Lucas BR, Hunter DJ. Targeting care: tailoring nonsurgical management according to clinical presentation. Rheum Dis Clin North Am. 2013;39(1):213-33.
39. Eyles JP, Lucas BR, Patterson JA, Williams MJ, Weeks K, Fransen M, et al. Does clinical presentation predict response to a nonsurgical chronic disease management program for endstage hip and knee osteoarthritis? J Rheumatol. 2014;41(11):2223-31.
40. Eyles JP, Mills K, Lucas BR, Williams MJ, Makovey J, Teoh L, et al. Can We Predict Those With Osteoarthritis Who Will Worsen Following a Chronic Disease Management Program? Arthritis Care Res (Hoboken). 2016;68(9):1268-77.
41. Eyles JP, Hunter DJ, Meneses SRF, Collins NJ, Dobson F, Lucas BR, et al. Instruments assessing attitudes toward or capability regarding self-management of osteoarthritis: a systematic review of measurement properties. Osteoarthritis Cartilage. 2017;25(8):1210-22.

42. Hibbard JH, Mahoney ER, Stockard J, Tusler M. Development and testing of a short form of the patient activation measure. Health Serv Res. 2005;40(6 Pt 1):1918-30.
43. Wagner EH. Chronic disease management: what will it take to improve care for chronic illness? Eff Clin Pract. 1998;1(1):2-4.
44. Epping-Jordan JE, Pruitt SD, Bengoa R, Wagner EH. Improving the quality of health care for chronic conditions. Qual Saf Health Care. 2004;13(4):299-305.
45. Health Change Australia. HealthChange [Available from: http://www.healthchange.com/.
46. Gale J, Skouteris H. Health coaching: Facilitating health behaviour change for chronic condition prevention and self-management. In: Caltabiano M, Ricciardelli L, editors. Applied Topics in Health Psychology Brisbane, Australia.: Wiley-Blackwell Publishers; 2013.
47. Agency for Clinical Innovation. Osteoarthritis Chronic Care Program Site Manual 2017 [cited 2018 15th August]. Available from: https://www.aci.health.nsw.gov.au/__data/assets/pdf_file/0006/390057/ACI_OACCP-Site-Manual-Oct-2017.pdf.
48. Tubach F, Ravaud P, Baron G, Falissard B, Logeart I, Bellamy N, et al. Evaluation of clinically relevant changes in patient reported outcomes in knee and hip osteoarthritis: the minimal clinically important improvement. Ann Rheum Dis. 2005;64(1):29-33.
49. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials. 1989;10(4):407-15.
50. Jordan JE, Haynes K, Livingston JA, Osborne RH. Comparison of the pre-post and transition question assessments in a health education setting. J Clin Epidemiol. 2009;62(6):642-9.
51. Terwee CB, Roorda LD, Dekker J, Bierma-Zeinstra SM, Peat G, Jordan KP, et al. Mind the MIC: large variation among populations and methods. J Clin Epidemiol. 2010;63(5):524-34.
52. Friedman L, Furberg C, DeMets D. Fundamentals of Clinical Trials. Fourth Edition ed. New York, USA: Springer; 2010.
53. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol. 1996;49(12):1373-9.
54. Field A. Discovering statistics using SPSS. 3rd ed. London, United Kingdom: Sage; 2009.
55. Austin PC, Steyerberg EW. The number of subjects per variable required in linear regression analyses. J Clin Epidemiol. 2015;68(6):627-36.
56. de Vet H, Terwee C, Mokkink L, Knol D. Measurement in Medicine: A Practical Guide to Biostatistics and Epidemiology. London: Cambridge University Press; 2011.
57. COSMIN. COSMIN helps you select the most suitable outcome measurement instruments 2018 [Available from: https://www.cosmin.nl/.
58. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63(7):737-45.

59. Terwee CB, Jansma EP, Riphagen, II, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual Life Res. 2009;18(8):1115-23.
60. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res. 2012;21(4):651-7.
61. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60(1):34-42.
62. Bond TG. Applying the Rasch model : fundamental measurement in the human sciences / authored by Trevor G. Bond and Christine M. Fox. Fox CM, Ebscohost, editors. New York: Routledge, Taylor & Francis Group; 2015.
63. Linacre JM. Winsteps® Rasch measurement computer program User's Guide. . Beaverton, Oregon: Winsteps.com2017.
64. World Health Organisation. Chronic diseases and health promotion: chronic rheumatic conditions [Available from: http://www.who.int/chp/topics/rheumatic/en/.
65. Murray C, Lopez A. The global burden of disease. A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020. Cambridge (MA): Harvard School of Public Health on behalf of the World Health Organization and The World Bank; 1996.
66. Centres for Disease Control and Prevention (CDC). National and state medical expenditures and lost earnings attributable to arthritis and other rheumatic conditions-United States. Morbidity and Mortality Weekly Report. 2007;56:4- 7.
67. Zhang W, Doherty M, Arden N, Bannwarth B, Bijlsma J, Gunther KP, et al. EULAR evidence based recommendations for the management of hip osteoarthritis: report of a task force of the EULAR Standing Committee for International Clinical Studies Including Therapeutics (ESCISIT). Ann Rheum Dis. 2005;64(5):669-81.
68. Zhang W, Nuki G, Moskowitz RW, Abramson S, Altman RD, Arden NK, et al. OARSI recommendations for the management of hip and knee osteoarthritis: part III: Changes in evidence following systematic cumulative update of research published through January 2009. Osteoarthritis Cartilage. 2010;18(4):476-99.
69. Pendleton A, Arden N, Dougados M, Doherty M, Bannwarth B, Bijlsma JW, et al. EULAR recommendations for the management of knee osteoarthritis: report of a task force of the Standing Committee for International Clinical Studies Including Therapeutic Trials (ESCISIT). Ann Rheum Dis. 2000;59(12):936-44.
70. Royal Australian College of General Practitioners. Guideline for the non-sugical management of hip and knee Osteoarthritis. 2009.
71. NICE and Royal College of Physicians Guidelines on Osteoarthritis. Osteoarthritis-national clinical guideline for care and management in adults. United Kingdom2008 [Available from: http://guidance.nice.org.uk/CG59.
72. Hochberg MC, Altman RD, April KT, Benkhalti M, Guyatt G, McGowan J, et al. American College of Rheumatology 2012 recommendations for the use of nonpharmacologic

and pharmacologic therapies in osteoarthritis of the hand, hip, and knee. Arthritis Care Res (Hoboken). 2012;64(4):465-74.
73. Misso ML, Pitt VJ, Jones KM, Barnes HN, Piterman L, Green SE. Quality and consistency of clinical practice guidelines for diagnosis and management of osteoarthritis of the hip and knee: a descriptive overview of published guidelines. Med J Aust. 2008;189(7):394-9.
74. Zhang W, Moskowitz RW, Nuki G, Abramson S, Altman RD, Arden N, et al. OARSI recommendations for the management of hip and knee osteoarthritis, Part II: OARSI evidence-based, expert consensus guidelines. Osteoarthritis Cartilage. 2008;16(2):137-62.
75. Janssen I, Mark AE. Separate and combined influence of body mass index and waist circumference on arthritis and knee osteoarthritis. Int J Obes (Lond). 2006;30(8):1223-8.
76. Chapple CM, Nicholson H, Baxter GD, Abbott JH. Patient characteristics that predict progression of knee osteoarthritis: a systematic review of prognostic studies. Arthritis Care Res (Hoboken). 2011;63(8):1115-25.
77. Heliovaara M, Makela M, Impivaara O, Knekt P, Aromaa A, Sievers K. Association of overweight, trauma and workload with coxarthrosis. A health survey of 7,217 persons. Acta Orthop Scand. 1993;64(5):513-8.
78. Karlson EW, Mandl LA, Aweh GN, Sangha O, Liang MH, Grodstein F. Total hip replacement due to osteoarthritis: The importance of age, obesity, and other modifiable risk factors. Am J Med. 2003;114(2):93-8.
79. Brand C, Hunter DJ, Hinman RS, March L, Osbourne R, Bennell KL. Improving care for people with OA of the hip and knee: how has national policy for OA been translated into service models in Australia? Int J Rheum Dis. 2011;14:181- 90.
80. Wolf S, Foley S, Budiman-Mak E, Moritz T, O'Connell S, Jelinek C, et al. Predictors of weight loss in overweight veterans with knee osteoarthritis who participated in a clinical trial. J Rehabil Res Dev. 2010;47(3):171-81.
81. Bingham CO, 3rd, Smugar SS, Wang H, Peloso PM, Gammaitoni A. Predictors of response to cyclo-oxygenase-2 inhibitors in osteoarthritis: pooled results from two identical trials comparing etoricoxib, celecoxib, and placebo. Pain Med. 2011;12(3):352-61.
82. Detora L, Krupta D, Bolognese J, Sperling R, Ehrich E. Rofecoxib shows consistent efficacy in OA clinical trials, regardless of specific patient demographic and disease factors. J Rheumatol. 2001;28(11):2494-503.
83. Robinson P, Keenan AM, Conaghan PG. Clinical effectiveness and dose response of image-guided intra-articular corticosteroid injection for hip osteoarthritis. Rheumatology (Oxford). 2007;46(2):285-91.
84. Kemper F, Gebhardt U, Meng T, Murray C. Tolerability and short-term effectiveness of hylan G-F 20 in 4253 patients with osteoarthritis of the knee in clinical practice. Curr Med Res Opin. 2005;21(8):1261-9.
85. Conrozier T, Mathieu P, Schott AM, Laurent I, Hajri T, Crozes P, et al. Factors predicting long-term efficacy of Hylan GF-20 viscosupplementation in knee osteoarthritis. Joint Bone Spine. 2003;70(2):128-33.
86. Anandacoomarasamy A, Bagga H, Ding CG, Burkhardt D, Sambrook PN, March LM. Predictors of clinical response to intraarticular hylan injections - A prospective study using

synovial fluid measures, clinical outcomes, and magnetic resonance imaging. J Rheumatol. 2008;35(4):685-90.
87. Goorman SD, Watanabe TK, Miller EH, Perry C. Functional outcome in knee osteoarthritis after treatment with hylan G-F 20: a prospective study. Arch Phys Med Rehabil. 2000;81(4):479-83.
88. Migliore A, Tormenta S, Massafra U, Bizzi E, Iannessi F, Alimonti A, et al. Intra-articular administration of hylan G-F 20 in patients with symptomatic hip osteoarthritis: tolerability and effectiveness in a large cohort study in clinical practice. Curr Med Res Opin. 2008;24(5):1309-16.
89. Bennett AN, Crossley KM, Brukner PD, Hinman RS. Predictors of symptomatic response to glucosamine in knee osteoarthritis: an exploratory study. Br J Sports Med. 2007;41(7):415-9.
90. Roseman T, Backenstrass M, Rosemann A, Szecsenyi J, Laux G. Predictors of depression in a sample of 1021 primary care patients within Osteoarthritis. Arthritis Rheum 2007;57(3):415–22.
91. Wise BL, Niu J, Zhang Y, Wang N, Jordan JM, Choy E, et al. Psychological factors and their relation to osteoarthritis pain. Osteoarthritis Cartilage. 2010;18(7):883-7.
92. Lin EB KW, Von Korff M, et al. . Effect of Improving Depression Care on Pain and Functional Outcomes Among Older Adults With Arthritis: A Randomized Controlled Trial. JAMA. 2003;290(18):2428- 9.
93. Jones A, Doherty M. Intra-articular corticosteroids are effective in osteoarthritis but there are no clinical predictors of response. Ann Rheum Dis. 1996;55(11):829-32.
94. Walker JS, Sheather-Reid RB, Carmody JJ, Vial JH, Day RO. Nonsteroidal antiinflammatory drugs in rheumatoid arthritis and osteoarthritis: support for the concept of "responders" and "nonresponders". Arthritis Rheum. 1997;40(11):1944-54.
95. Palmieri-Smith RM, Thomas AC, Karvonen-Gutierrez C, Sowers MF. Isometric quadriceps strength in women with mild, moderate, and severe knee osteoarthritis. Am J Phys Med Rehabil. 2010;89(7):541-8.
96. Segal NA, Torner JC, Felson D, Niu J, Sharma L, Lewis CE, et al. Effect of thigh strength on incident radiographic and symptomatic knee osteoarthritis in a longitudinal cohort. Arthritis Rheum. 2009;61(9):1210-7.
97. Amin S, Baker K, Niu J, Clancy M, Goggins J, Guermazi A, et al. Quadriceps strength and the risk of cartilage loss and symptom progression in knee osteoarthritis. Arthritis Rheum. 2009;60(1):189-98.
98. Roos EM, Herzog W, Block JA, Bennell KL. Muscle weakness, afferent sensory dysfunction and exercise in knee osteoarthritis. Nat Rev Rheumatol. 2011;7(1):57-63.
99. Suetta C, Aagaard P, Magnusson SP, Andersen LL, Sipila S, Rosted A, et al. Muscle size, neuromuscular activation, and rapid force characteristics in elderly men and women: effects of unilateral long-term disuse due to hip-osteoarthritis. J Appl Physiol. 2007;102(3):942-8.
100. Fransen M, McConnell S. Land-based exercise for osteoarthritis of the knee: a metaanalysis of randomized controlled trials. J Rheumatol. 2009;36(6):1109-17.

101. Fransen M, McConnell S, Hernandez-Molina G, Reichenbach S. Does land-based exercise reduce pain and disability associated with hip osteoarthritis? A meta-analysis of randomized controlled trials. Osteoarthritis Cartilage. 2010;18(5):613-20.
102. Wright AA, Cook CE, Flynn TW, Baxter GD, Abbott JH. Predictors of response to physical therapy intervention in patients with primary hip osteoarthritis. Phys Ther. 2011;91(4):510-24.
103. Sattler M, Dannhauer T, Hudelmaier M, Wirth W, Sanger AM, Kwoh CK, et al. Side differences of thigh muscle cross-sectional areas and maximal isometric muscle force in bilateral knees with the same radiographic disease stage, but unilateral frequent pain - data from the osteoarthritis initiative. Osteoarthritis Cartilage. 2012;20(6):532-40.
104. Pietrosimone BG, Hertel J, Ingersoll CD, Hart JM, Saliba SA. Voluntary quadriceps activation deficits in patients with tibiofemoral osteoarthritis: a meta-analysis. PM R. 2011;3(2):153-62; quiz 62.
105. Scopaz KA, Piva SR, Gil AB, Woollard JD, Oddis CV, Fitzgerald GK. Effect of baseline quadriceps activation on changes in quadriceps strength after exercise therapy in subjects with knee osteoarthritis. Arthritis Rheum. 2009;61(7):951-7.
106. Gudbergsen H, Boesen M, Lohmander LS, Christensen R, Henriksen M, Bartels EM, et al. Weight loss is effective for symptomatic relief in obese subjects with knee osteoarthritis independently of joint damage severity assessed by high-field MRI and radiography. Osteoarthritis Cartilage. 2012;20(6):495-502.
107. Hunter DJ, Niu J, Felson DT, Harvey WF, Gross KD, McCree P, et al. Knee alignment does not predict incident osteoarthritis: the Framingham osteoarthritis study. Arthritis Rheum. 2007;56(4):1212-8.
108. Brouwer GM, van Tol AW, Bergink AP, Belo JN, Bernsen RMD, Reijman M, et al. Association between valgus and varus alignment and the development and progression of radiographic osteoarthritis of the knee. Arthritis Rheum. 2007;56(4):1204-11.
109. Hinman RS, Payne C, Metcalf BR, Wrigley TV, Bennell KL. Lateral wedges in knee osteoarthritis: what are their immediate clinical and biomechanical effects and can these predict a three-month clinical outcome? Arthritis Rheum. 2008;59(3):408-15.
110. Derek T, Cooke V, Sled EA, Scudamore RA. Frontal plane knee alignment: A call for standardized measurement. J Rheumatol. 2007;34(9):1796-801.
111. Andriacchi TP. Dynamics of knee malalignment. Orthop Clin North Am. 1994;25(3):395-403.
112. Bennell KL, Hunt MA, Wrigley TV, Hunter DJ, McManus FJ, Hodges PW, et al. Hip strengthening reduces symptoms but not knee load in people with medial knee osteoarthritis and varus malalignment: a randomised controlled trial. Osteoarthritis Cartilage. 2010;18(5):621-8.
113. Foroughi N, Smith RM, Lange AK, Singh MA, Vanwanseele B. Progressive resistance training and dynamic alignment in osteoarthritis: A single-blind randomised controlled trial. Clin Biomech (Bristol, Avon). 2011;26(1):71-7.
114. Sharma L, Dunlop DD, Cahue S, Song J, Hayes KW. Quadriceps strength and osteoarthritis progression in malaligned and lax knees. Ann Intern Med. 2003;138(8):613-9.

115. Sharma L, Song J, Felson DT, Cahue S, Shamiyeh E, Dunlop DD. The role of knee alignment in disease progression and functional decline in knee osteoarthritis. JAMA. 2001;286(2):188-95.
116. Lim BW, Hinman RS, Wrigley TV, Bennell KL. Varus malalignment and its association with impairments and functional limitations in medial knee osteoarthritis. Arthritis Rheum. 2008;59(7):935-42.
117. Lim B-W, Hinman RS, Wrigley TV, Sharma L, Bennell KL. Does knee malalignment mediate the effects of quadriceps strengthening on knee adduction moment, pain and function in medial knee osteoarthritis? A randomized controlled trial. Arthritis Rheum. 2008;59(7):943-51.
118. Guermazi A, Burstein D, Conaghan P, Eckstein F, Hellio Le Graverand-Gastineau MP, Keen H, et al. Imaging in osteoarthritis. Rheum Dis Clin North Am. 2008;34(3):645-87.
119. Conaghan PG, Hunter DJ, Maillefert JF, Reichmann WM, Losina E. Summary and recommendations of the OARSI FDA osteoarthritis Assessment of Structural Change Working Group. Osteoarthritis Cartilage. 2011;19(5):606-10.
120. Gudbergsen H, Boesen M, Christensen R, Astrup A, Bliddal H. Radiographs and low field MRI (0.2T) as predictors of efficacy in a weight loss trial in obese women with knee osteoarthritis. BMC Musculoskelet Disord. 2011;12:56.
121. Case JP, Baliunas AJ, Block JA. Lack of efficacy of acetaminophen in treating symptomatic knee osteoarthritis: a randomized, double-blind, placebo-controlled comparison trial with diclofenac sodium. Arch Intern Med. 2003;163(2):169-78.
122. Deshmukh AJ, Panagopoulos G, Alizadeh A, Rodriguez JA, Klein DA. Intra-articular hip injection: does pain relief correlate with radiographic severity of osteoarthritis? Skeletal Radiol. 2011;40(11):1449-54.
123. Plant MJ, Borg AA, Dziedzic K, Saklatvala J, Dawes PT. Radiographic patterns and response to corticosteroid hip injection. Ann Rheum Dis. 1997;56(8):476-80.
124. Atchia I, Kane D, Reed MR, Isaacs JD, Birrell F. Efficacy of a single ultrasound-guided injection for the treatment of hip osteoarthritis. Ann Rheum Dis. 2011;70(1):110-6.
125. Loeser RF, Goldring SR, Scanzello CR, Goldring MB. Osteoarthritis: a disease of the joint as an organ. Arthritis Rheum. 2012;64(6):1697-707.
126. Chao J, Wu C, Sun B, Hose MK, Quan A, Hughes TH, et al. Inflammatory characteristics on ultrasound predict poorer longterm response to intraarticular corticosteroid injections in knee osteoarthritis. J Rheumatol. 2010;37(3):650-5.
127. Pendleton A, Millar A, O'Kane D, Wright GD, Taggart AJ. Can sonography be used to predict the response to intra-articular corticosteroid injection in primary osteoarthritis of the knee? Scand J Rheumatol. 2008;37(5):395-7.
128. Babyak MA. What you see may not be what you get: A brief, nontechnical introduction to overfitting in regression-type models. Psychosomatic Medicine. 2004;66(3):411-21.
129. French HP, Galvin R, Cusack T, McCarthy GM. Predictors of short-term outcome to exercise and manual therapy for people with hip osteoarthritis. Physical Therapy. 2014;94(1):31-9.

130. Skou ST, Simonsen ME, Odgaard A, Roos EM. Predictors of long-term effect from education and exercise in patients with knee and hip pain. Dan Med J. 2014;61(7):A4867.
131. Taylor SS, Oddone EZ, Coffman CJ, Jeffreys AS, Bosworth HB, Allen KD. Cognitive Mediators of Change in Physical Functioning in Response to a Multifaceted Intervention for Managing Osteoarthritis. Intl J Behav Med. 2018;25(2):162-70.
132. Hofstede SN, Marang-van de Mheen PJ, Vliet Vlieland TP, van den Ende CH, Nelissen RG, van Bodegom-Vos L. Barriers and Facilitators Associated with Non-Surgical Treatment Use for Osteoarthritis Patients in Orthopaedic Practice. PLoS One. 2016;11(1):e0147406.
133. Jordan K, Clarke AM, Symmons DP, Fleming D, Porcheret M, Kadam UT, et al. Measuring disease prevalence: a comparison of musculoskeletal disease using four general practice consultation databases. Br J Gen Pract. 2007;57(534):7-14.
134. Agency for Clinical Innovation. Musculoskeletal Network Osteoarthritis Chronic Care Program Model of Care 2018 [Available from: https://www.aci.health.nsw.gov.au/resources/musculoskeletal/osteoarthritis_chronic_care_program/osteoarthritis-chronic-care-program.
135. Thorborg K, Roos EM, Bartels EM, Petersen J, Holmich P. Validity, reliability and responsiveness of patient-reported outcome questionnaires when assessing hip and groin disability: a systematic review. Br J Sports Med. 2010;44(16):1186-96.
136. McPoil TG, Vicenzino B, Cornwall MW, Collins N, Warren M. Reliability and normative values for the foot mobility magnitude: a composite measure of vertical and medial-lateral mobility of the midfoot. J Foot Ankle Res. 2009;2(1):6.
137. KOOS User's Guide 2012 [cited 2018. Available from: http://www.koos.nu/.
138. HOOS User's Guide 2003 [Available from: www.koos.nu/HOOSGuide2003.pdf.
139. Gloster AT, Rhoades HM, Novy D, Klotsche J, Senior A, Kunik M, et al. Psychometric properties of the Depression Anxiety and Stress Scale-21 in older primary care patients. J Affect Disord. 2008;110(3):248-59.
140. Lovibond SH, Lovibond PF. Manual for the Depression Anxiety Stress Scales. . 2nd ed. Sydney: Psychology Foundation; 1995.
141. Sangha O, Stucki G, Liang MH, Fossel AH, Katz JN. The Self-Administered Comorbidity Questionnaire: a new method to assess comorbidity for clinical and health services research. Arthritis Rheum. 2003;49(2):156-63.
142. Dobson F, Hinman RS, Roos EM, Abbott JH, Stratford P, Davis AM, et al. OARSI recommended performance-based tests to assess physical function in people diagnosed with hip or knee osteoarthritis. Osteoarthritis Cartilage. 2013;21(8):1042-52.
143. Bennell K, Dobson F, Hinman R. Measures of physical performance assessments: Self-Paced Walk Test (SPWT), Stair Climb Test (SCT), Six-Minute Walk Test (6MWT), Chair Stand Test (CST), Timed Up & Go (TUG), Sock Test, Lift and Carry Test (LCT), and Car Task. Arthritis Care Res (Hoboken).63 Suppl 11:S350-70.
144. Peduzzi P, Concato J, Holford T, Feinstein A. The importance of events per independent variable in multivariable analysis, II: accuracy and precision of regression estimates. . J Clin Epidemiol 1995;48(12):1503-10.

145. Dobson F, Hinman RS, Hall M, Terwee CB, Roos EM, Bennell KL. Measurement properties of performance-based measures to assess physical function in hip and knee osteoarthritis: a systematic review. Osteoarthritis Cartilage. 2012;20(12):1548-62.
146. Felson DT, Zhang Y, Hannan MT, Naimark A, Weissman BN, Aliabadi P, et al. The incidence and natural history of knee osteoarthritis in the elderly. The Framingham Osteoarthritis Study. Arthritis Rheum. 1995;38(10):1500-5.
147. Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet. 2012;380(9859):2163-96.
148. Larmer PJ, Reay ND, Aubert ER, Kersten P. Systematic review of guidelines for the physical management of osteoarthritis. Arch Phys Med Rehabil. 2014;95(2):375-89.
149. Zhang W, Moskowitz RW, Nuki G, Abramson S, Altman RD, Arden N, et al. OARSI recommendations for the management of hip and knee osteoarthritis, part I: critical appraisal of existing treatment guidelines and systematic review of current research evidence. Osteoarthritis Cartilage. 2007;15(9):981-1000.
150. Raynauld JP, Martel-Pelletier J, Haraoui B, Choquette D, Dorais M, Wildi LM, et al. Risk factors predictive of joint replacement in a 2-year multicentre clinical trial in knee osteoarthritis using MRI: results from over 6 years of observation. Ann Rheum Dis. 2011;70(8):1382-8.
151. Conaghan PG, D'Agostino MA, Le Bars M, Baron G, Schmidely N, Wakefield R, et al. Clinical and ultrasonographic predictors of joint replacement for knee osteoarthritis: results from a large, 3-year, prospective EULAR study. Ann Rheum Dis. 2010;69(4):644-7.
152. Hawker GA, Guan J, Croxford R, Coyte PC, Glazier RH, Harvey BJ, et al. A prospective population-based study of the predictors of undergoing total joint arthroplasty. Arthritis Rheum. 2006;54(10):3212-20.
153. Wright AA, Cook C, Abbott JH. Variables associated with the progression of hip osteoarthritis: a systematic review. Arthritis Rheum. 2009;61(7):925-36.
154. Gossec L, Paternotte S, Bingham CO, 3rd, Clegg DO, Coste P, Conaghan PG, et al. OARSI/OMERACT initiative to define states of severity and indication for joint replacement in hip and knee osteoarthritis. An OMERACT 10 Special Interest Group. J Rheumatol. 2011;38(8):1765-9.
155. Manno RL, Bingham CO, 3rd, Paternotte S, Gossec L, Halhol H, Giacovelli G, et al. OARSI-OMERACT initiative: defining thresholds for symptomatic severity and structural changes in disease modifying osteoarthritis drug (DMOAD) clinical trials. Osteoarthritis Cartilage. 2012;20(2):93-101.
156. Ackerman IN, Bennell KL, Osborne RH. Decline in Health-Related Quality of Life reported by more than half of those waiting for joint replacement surgery: a prospective cohort study. BMC Musculoskelet Disord. 2011;12:108.
157. Hawker GA, Mian S, Kendzerska T, French M. Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of

Intermittent and Constant Osteoarthritis Pain (ICOAP). Arthritis Care Res (Hoboken). 2011;63 Suppl 11:S240-52.
158. Roos EM, Lohmander LS. The Knee injury and Osteoarthritis Outcome Score (KOOS): from joint injury to osteoarthritis. Health Qual Life Outcomes. 2003;1:64.
159. Lovibond PF, Lovibond SH. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. Behav Res Ther. 1995;33(3):335-43.
160. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR, Clinical Significance Consensus Meeting G. Methods to explain the clinical significance of health status measures. Mayo Clin Proc. 2002;77(4):371-83.
161. Tabachnick BG, Fidell LS. Using Multivariate Statistics. 5 ed. New York: Allyn & Bacon; 2006.
162. Naylor JM, Hayen A, Davidson E, Hackett D, Harris IA, Kamalasena G, et al. Minimal detectable change for mobility and patient-reported tools in people with osteoarthritis awaiting arthroplasty. BMC Musculoskelet Disord. 2014;15:235.
163. Hoogeboom TJ, van den Ende CH, van der Sluis G, Elings J, Dronkers JJ, Aiken AB, et al. The impact of waiting for total joint replacement on pain and functional status: a systematic review. Osteoarthritis Cartilage. 2009;17(11):1420-7.
164. Fowler FJ, Jr., Gerstein BS, Barry MJ. How patient centered are medical decisions?: Results of a national survey. JAMA Intern Med. 2013;173(13):1215-21.
165. Sepucha K, Feibelmann S, Chang Y, Clay CF, Kearing SA, Tomek I, et al. Factors associated with the quality of patients' surgical decisions for treatment of hip and knee osteoarthritis. J Am Coll Surg. 2013;217(4):694-701.
166. Escobar A, Gonzalez M, Quintana JM, Vrotsou K, Bilbao A, Herrera-Espineira C, et al. Patient acceptable symptom state and OMERACT-OARSI set of responder criteria in joint replacement. Identification of cut-off values. Osteoarthritis Cartilage. 2012;20(2):87-92.
167. Grenard JL, Munjas BA, Adams JL, Suttorp M, Maglione M, McGlynn EA, et al. Depression and medication adherence in the treatment of chronic diseases in the United States: a meta-analysis. J Gen Intern Med. 2011;26(10):1175-82.
168. Kapstad H, Rustoen T, Hanestad BR, Moum T, Langeland N, Stavem K. Changes in pain, stiffness and physical function in patients with osteoarthritis waiting for hip or knee joint replacement surgery. Osteoarthritis Cartilage. 2007;15(7):837-43.
169. Gandhi R, Wasserstein D, Razak F, Davey JR, Mahomed NN. BMI independently predicts younger age at hip and knee replacement. Obesity (Silver Spring). 2010;18(12):2362-6.
170. Riddle DL, Stratford PW. Body weight changes and corresponding changes in pain and function in persons with symptomatic knee osteoarthritis: a cohort study. Arthritis Care Res (Hoboken). 2013;65(1):15-22.
171. Tuominen U, Blom M, Hirvonen J, Seitsalo S, Lehto M, Paavolainen P, et al. The effect of co-morbidities on health-related quality of life in patients placed on the waiting list for total joint replacement. Health Qual Life Outcomes. 2007;5:16.

172. Coulter A, Entwistle VA, Eccles A, Ryan S, Shepperd S, Perera R. Personalised care planning for adults with chronic or long-term health conditions. Cochrane Database Syst Rev. 2015(3):CD010523.
173. Moher D, Liberati A, Tetzlaff J, Altman D, PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. Ann Int Med. 2009;151(4):264-9.
174. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res. 2010;19(4):539-49.
175. Ahn YH, Yi CH, Ham OK, Kim BJ. Psychometric Properties of the Korean Version of the "Patient Activation Measure 13" (PAM13-K) in Patients With Osteoarthritis. Eval Health Prof. 2015;38(2):255-64.
176. Heuts PH, de Bie RA, Dijkstra A, Aretz K, Vlaeyen JW, Schouten HJ, et al. Assessment of readiness to change in patients with osteoarthritis. development and application of a new questionnaire. Clin Rehabil. 2005;19(3):290-9.
177. Kelly PA, Kallen MA, Suarez-Almazor ME. A combined-method psychometric analysis recommended modification of the multidimensional health locus of control scales. J Clin Epidemiol. 2007;60(5):440-7.
178. Liu Y, Doucette WR, Farris KB. Perceived difficulty and self-efficacy in the factor structure of perceived behavioral control to seek drug information from physicians and pharmacists. Res Social Adm Pharm. 2007;3(2):145-59.
179. Ndosi M, Bremander A, Hamnes B, Horton M, Kukkurainen ML, Machado P, et al. Validation of the educational needs assessment tool as a generic instrument for rheumatic diseases in seven European countries. Ann Rheum Dis. 2014;73(12):2122-9.
180. ten Klooster PM, Oostveen JC, Zandbelt LC, Taal E, Drossaert CH, Harmsen EJ, et al. Further validation of the 5-item Perceived Efficacy in Patient-Physician Interactions (PEPPI-5) scale in patients with osteoarthritis. Patient Educ Couns. 2012;87(1):125-30.
181. ten Klooster PM, Taal E, Siemons L, Oostveen JC, Harmsen EJ, Tugwell PS, et al. Translation and validation of the Dutch version of the Effective Consumer Scale (EC-17). Qual Life Res. 2013;22(2):423-9.
182. Zhao H, Luo W, Maly RC, Liu J, Lee J, Cui Y. Validation of the Chinese version 10-item Perceived Efficacy in Patient-Physician Interactions scale in patients with osteoarthritis. Patient Prefer Adherence. 2016;10:2189-95.
183. Hibbard JH, Stockard J, Mahoney ER, Tusler M. Development of the Patient Activation Measure (PAM): conceptualizing and measuring activation in patients and consumers. Health Serv Res. 2004;39(4 Pt 1):1005-26.
184. Kristjansson E, Tugwell PS, Wilson AJ, Brooks PM, Driedger SM, Gallois C, et al. Development of the effective musculoskeletal consumer scale. J Rheumatol. 2007;34(6):1392-400.
185. Maly RC, Frank JC, Marshall GN, DiMatteo MR, Reuben DB. Perceived efficacy in patient-physician interactions (PEPPI): validation of an instrument in older persons. J Am Geriatr Soc. 1998;46(7):889-94.

186. Skou ST, Odgaard A, Rasmussen JO, Roos EM. Group education and exercise is feasible in knee and hip osteoarthritis. Dan Med J. 2012;59(12):A4554.
187. Garratt AM, Lochting I, Smedslund G, Hagen KB. Measurement properties of instruments assessing self-efficacy in patients with rheumatic diseases. Rheumatology (Oxford). 2014;53(7):1161-71.
188. Bandura A. Self-Efficacy. The Corsini Encyclopedia of Psychology: John Wiley & Sons, Inc.; 2010.
189. Osborne RH, Elsworth GR, Whitfield K. The Health Education Impact Questionnaire (heiQ): an outcomes and evaluation measure for patient education and self-management interventions for people with chronic conditions. Patient Educ Couns. 2007;66(2):192-201.
190. Umapathy H, Bennell K, Dickson C, Dobson F, Fransen M, Jones G, et al. The Web-Based Osteoarthritis Management Resource My Joint Pain Improves Quality of Care: A Quasi-Experimental Study. J Med Internet Res. 2015;17(7):e167.
191. Lorig K, Chastain RL, Ung E, Shoor S, Holman HR. Development and evaluation of a scale to measure perceived self-efficacy in people with arthritis. Arthritis Rheum. 1989;32(1):37-44.
192. Thorstensson CA, Garellick G, Rystedt H, Dahlberg LE. Better Management of Patients with Osteoarthritis: Development and Nationwide Implementation of an Evidence-Based Supported Osteoarthritis Self-Management Programme. Musculoskeletal care. 2015;13(2):67-75.
193. McAllister M, Dunn G, Payne K, Davies L, Todd C. Patient empowerment: the need to consider it as a measurable patient-reported outcome for chronic conditions. BMC Health Serv Res. 2012;12:157.
194. Moljord IE, Lara-Cabrera ML, Perestelo-Perez L, Rivero-Santana A, Eriksen L, Linaker OM. Psychometric properties of the Patient Activation Measure-13 among out-patients waiting for mental health treatment: A validation study in Norway. Patient Educ Couns. 2015;98(11):1410-7.
195. Packer TL, Kephart G, Ghahari S, Audulv A, Versnel J, Warner G. The Patient Activation Measure: a validation study in a neurological population. Qual Life Res. 2015;24(7):1587-96.
196. Schmaderer M, Pozehl B, Hertzog M, Zimmerman L. Psychometric Properties of the Patient Activation Measure in Multimorbid Hospitalized Patients. J Nurs Meas. 2015;23(3):128-41.
197. Skolasky RL, Green AF, Scharfstein D, Boult C, Reider L, Wegener ST. Psychometric properties of the patient activation measure among multimorbid older adults. Health Serv Res. 2011;46(2):457-78.
198. Santesso N, Rader T, Wells GA, O'Connor AM, Brooks PM, Driedger M, et al. Responsiveness of the Effective Consumer Scale (EC-17). J Rheumatol. 2009;36(9):2087-91.
199. Hibbard JH, Greene J, Shi Y, Mittler J, Scanlon D. Taking the long view: how well do patient activation scores predict outcomes four years later? Med Care Res Rev. 2015;72(3):324-37.
200. Hibbard JH, Mahoney ER, Stock R, Tusler M. Do increases in patient activation result in improved self-management behaviors? Health Serv Res. 2007;42(4):1443-63.

201. Sacks RM, Greene J, Hibbard J, Overton V, Parrotta CD. Does patient activation predict the course of type 2 diabetes? A longitudinal study. Patient Educ Couns. 2017;100(7):1268-75.
202. Rasch G. Probabilistic model for some intelligence and achievement tests. Denmark: Danish Institute for Educational Research.; 1960.
203. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis Rheum. 2007;57(8):1358-62.
204. Richardson J, Atherton Day N, Peacock S, Iezzi A. Measurement of the Quality of Life for Economic Evaluation and the Assessment of Quality of Life (AQoL) Mark 2 Instrument. The Australian Economic Review. 2004;37(1):62-88.
205. Linacre J. Sample Size and Item Calibration Stability. Rasch Measurement Transactions. 1994;7(4):328.
206. Christensen KB, Makransky G, Horton M. Critical Values for Yen's Q3: Identification of Local Dependence in the Rasch Model Using Residual Correlations. Appl Psychol Meas. 2017;41(3):178-94.
207. Blakemore A, Hann M, Howells K, Panagioti M, Sidaway M, Reeves D, et al. Patient activation in older people with long-term conditions and multimorbidity: correlates and change in a cohort study in the United Kingdom. BMC Health Serv Res. 2016;16(1):582.
208. Magnezi R, Glasser S, Shalev H, Sheiber A, Reuveni H. Patient activation, depression and quality of life. Patient Educ Couns. 2014;94(3):432-7.
209. Stepleman L, Rutter MC, Hibbard J, Johns L, Wright D, Hughes M. Validation of the patient activation measure in a multiple sclerosis clinic sample and implications for care. Disabil Rehabil. 2010;32(19):1558-67.
210. Cohen J. Statistical power analysis for the behavioral sciences (2nd ed.). Erlbaum. L, editor. New Jersey1988.
211. Insignia Health. Patient Activation Measure (PAM) 13™. Portland, Oregon; Insignia Health LLC; 2014.
212. Graffigna G, Barello S, Bonanomi A, Lozza E, Hibbard J. Measuring patient activation in Italy: Translation, adaptation and validation of the Italian version of the patient activation measure 13 (PAM13-I). BMC Med Inform Decis Mak. 2015;15:109.
213. Maindal HT, Sokolowski I, Vedsted P. Translation, adaptation and validation of the American short form Patient Activation Measure (PAM13) in a Danish version. BMC Public Health. 2009;9:209.
214. Zill JM, Dwinger S, Kriston L, Rohenkohl A, Harter M, Dirmaier J. Psychometric evaluation of the German version of the Patient Activation Measure (PAM13). BMC Public Health. 2013;13:1027.
215. Green CA, Perrin NA, Polen MR, Leo MC, Hibbard JH, Tusler M. Development of the Patient Activation Measure for mental health. Adm Policy Ment Health. 2010;37(4):327-33.
216. Harrison C, Henderson J, Miller G, Britt H. The prevalence of complex multimorbidity in Australia. Aust N Z J Public Health. 2016;40(3):239-44.

217. Scott NW, Fayers PM, Aaronson NK, Bottomley A, de Graeff A, Groenvold M, et al. A simulation study provided sample size guidance for differential item functioning (DIF) studies using short scales. J Clin Epidemiol. 2009;62(3):288-95.
218. Teresi JA. Different approaches to differential item functioning in health applications. Advantages, disadvantages and some neglected topics. Med Care. 2006;44(11 Suppl 3):S152-70.
219. Roberts NJ, Kidd L, Dougall N, Patel IS, McNarry S, Nixon C. Measuring patient activation: The utility of the Patient Activation Measure within a UK context-Results from four exemplar studies and potential future applications. Patient Educ Couns. 2016;99(10):1739-46.
220. Osteras N, Jordan KP, Clausen B, Cordeiro C, Dziedzic K, Edwards J, et al. Self-reported quality care for knee osteoarthritis: comparisons across Denmark, Norway, Portugal and the UK. RMD Open. 2015;1(1):e000136.
221. Dobson F, Bennell KL, French SD, Nicolson PJ, Klaasman RN, Holden MA, et al. Barriers and Facilitators to Exercise Participation in People with Hip and/or Knee Osteoarthritis: Synthesis of the Literature Using Behavior Change Theory. Am J Phys Med Rehabil. 2016;95(5):372-89.
222. Burgess E, Hassmen P, Pumpa KL. Determinants of adherence to lifestyle intervention in adults with obesity: a systematic review. Clin Obes. 2017;7(3):123-35.
223. Skolasky RL, Mackenzie EJ, Wegener ST, Riley LH, 3rd. Patient activation and adherence to physical therapy in persons undergoing spine surgery. Spine (Phila Pa 1976). 2008;33(21):E784-91.
224. Gandek B. Measurement properties of the Western Ontario and McMaster Universities Osteoarthritis Index: a systematic review. Arthritis Care Res (Hoboken). 2015;67(2):216-29.
225. Becker NV, Asch DA, Kullgren JT, Bellamy SL, Sen AP, Volpp KG. Stages of change and patient activation measure scores in the context of incentive-based health interventions. Am J Health Promot. 2015;30(2):133-5.
226. Andrawis J, Akhavan S, Chan V, Lehil M, Pong D, Bozic KJ. Higher Preoperative Patient Activation Associated With Better Patient-reported Outcomes After Total Joint Arthroplasty. Clin Orthop Relat Res. 2015;473(8):2688-97.
227. Skou ST, Roos EM, Laursen MB, Rathleff MS, Arendt-Nielsen L, Simonsen O, et al. A Randomized, Controlled Trial of Total Knee Replacement. N Engl J Med. 2015;373(17):1597-606.
228. Wilkie R, Blagojevic-Bucknall M, Jordan KP, Pransky G. Onset of work restriction in employed adults with lower limb joint pain: individual factors and area-level socioeconomic conditions. J Occup Rehabil. 2013;23(2):180-8.
229. Ackerman IN, Ademi Z, Osborne RH, Liew D. Comparison of health-related quality of life, work status, and health care utilization and costs according to hip and knee joint disease severity: a national Australian study. Phys Ther. 2013;93(7):889-99.
230. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev. 1977;84(2):191-215.

231. Ellis MI, Seedhom BB, Wright V. Forces in the knee joint whilst rising from a seated position. J Biomed Eng. 1984;6(2):113-20.
232. Knoop J, van der Leeden M, Thorstensson CA, Roorda LD, Lems WF, Knol DL, et al. Identification of phenotypes with different clinical outcomes in knee osteoarthritis: data from the Osteoarthritis Initiative. Arthritis Care Res (Hoboken). 2011;63(11):1535-42.
233. Lin EH, Tang L, Katon W, Hegel MT, Sullivan MD, Unutzer J. Arthritis pain and disability: response to collaborative depression care. Gen Hosp Psychiatry. 2006;28(6):482-6.
234. Mallen CD, Nicholl BI, Lewis M, Bartlam B, Green D, Jowett S, et al. The effects of implementing a point-of-care electronic template to prompt routine anxiety and depression screening in patients consulting for osteoarthritis (the Primary Care Osteoarthritis Trial): A cluster randomised trial in primary care. PLoS Med. 2017;14(4):e1002273.
235. Lau B, Gange SJ, Moore RD. Interval and clinical cohort studies: epidemiological issues. AIDS Res Hum Retroviruses. 2007;23(6):769-76.
236. Allen KD, Bosworth HB, Chatterjee R, Coffman CJ, Corsino L, Jeffreys AS, et al. Clinic variation in recruitment metrics, patient characteristics and treatment use in a randomized clinical trial of osteoarthritis management. BMC Musculoskelet Disord. 2014;15:413.
237. Zhang Y, Jordan JM. Epidemiology of osteoarthritis. Clin Geriatr Med. 2010;26(3):355-69.
238. Bierma-Zeinstra SM, Verhagen AP. Osteoarthritis subpopulations and implications for clinical trial design. Arthritis Res Ther. 2011;13(2):213.
239. Stanton TR, Hancock MJ, Maher CG, Koes BW. Critical appraisal of clinical prediction rules that aim to optimize treatment selection for musculoskeletal conditions. Phys Ther. 2010;90(6):843-54.
240. Mobasheri A, Batt M. An update on the pathophysiology of osteoarthritis. Ann Phys Rehabil Med. 2016;59(5-6):333-9.
241. Bruyere O, Cooper C, Arden N, Branco J, Brandi ML, Herrero-Beaumont G, et al. Can we identify patients with high risk of osteoarthritis progression who will respond to treatment? A focus on epidemiology and phenotype of osteoarthritis. Drugs Aging. 2015;32(3):179-87.
242. Deveza LA, Loeser RF. Is osteoarthritis one disease or a collection of many? Rheumatology (Oxford). 2018;57(suppl_4):iv34-iv42.
243. Karsdal MA, Christiansen C, Ladel C, Henriksen K, Kraus VB, Bay-Jensen AC. Osteoarthritis--a case for personalized health care? Osteoarthritis Cartilage. 2014;22(1):7-16.
244. Deveza LA, Melo L, Yamato TP, Mills K, Ravi V, Hunter DJ. Knee osteoarthritis phenotypes and their relevance for outcomes: a systematic review. Osteoarthritis Cartilage. 2017;25(12):1926-41.

Appendices

Appendix 1. Search strategy

i) **Construct**

generalized self efficacy scale[tiab] OR adaptive behavior[tiab] OR multidimensional health locus of control[tiab] OR pain self efficacy questionnaire[tiab] OR health literacy management scale[tiab] OR stages of change questionnaire in osteoarthritis[tiab] OR health education impact questionnaire[tiab] OR patient activation measure[tiab] OR effective consumer scale[tiab] OR arthritis self-efficacy scale[tiab] OR internal-external control[MH] OR locus of control[tw] OR attitude to health[MH] OR health locus of control[tiab] OR adaptation, psychological[MH] OR health behavior[MH] OR health knowledge, attitudes, practice[MH] OR self management behavio*[tiab] OR patient activation[tiab] OR self concept[MH] OR self efficacy[MH] OR confidence[tiab] OR activation[tiab] OR consumer participation[MH] OR patient education as topic[MH] OR Patient Participation[MH] OR individualized medicine[MH] OR patient-centered care[MH] OR goals[MH] OR patient preference[MH] OR choice behavior[MH] OR decision making[MH] OR patient care planning[MH] OR personalised care planning[tiab] OR patient led[tiab] OR selftreatment[tiab] OR self treat*[tiab] OR self manage*[tiab] OR self care[tiab] OR self care[MH]

ii) Target population

osteoarthritis[MH] OR osteoarth*[tiab] OR degenerative arthritis[tiab] OR arthrosis[tiab]

iii) Measurement instrument filter

instrumentation[sh] OR methods[sh] OR validation studies[pt] OR Comparative Study[pt] OR psychometrics[MH] OR psychometr*[tiab] OR clinimetr*[tw] OR clinometr*[tw] OR "outcome assessment (health care)"[MH] OR "outcome assessment"[tiab] OR "outcome measure*"[tw] OR "observer variation"[MH] OR "observer variation"[tiab] OR "Health Status Indicators"[MH] OR "reproducibility of results"[MH] OR reproducib*[tiab] OR "discriminant analysis"[MH] OR reliab*[tiab] OR unreliab*[tiab] OR valid*[tiab] OR coefficient[tiab] OR homogeneity[tiab] OR homogeneous[tiab] OR "internal consistency"[tiab] OR (cronbach*[tiab] AND (alpha[tiab] OR alphas[tiab])) OR (item[tiab] AND (correlation*[tiab] OR selection*[tiab] OR reduction*[tiab])) OR agreement[tiab] OR precision[tiab] OR

imprecision[tiab] OR "precise values"[tiab] OR test–retest[tiab] OR (test[tiab] AND retest[tiab]) OR (reliab*[tiab] AND (test[tiab] OR retest[tiab])) OR stability[tiab] OR interrater[tiab] OR inter-rater[tiab] OR intrarater[tiab] OR intra-rater[tiab] OR intertester[tiab] OR inter-tester[tiab] OR intratester[tiab] OR intra-tester[tiab] OR interobserver[tiab] OR inter-observer[tiab] OR intraobserver[tiab] OR intraobserver[tiab] OR intertechnician[tiab] OR inter-technician[tiab] OR intratechnician[tiab] OR intra-technician[tiab] OR interexaminer[tiab] OR inter-examiner[tiab] OR intraexaminer[tiab] OR intra-examiner[tiab] OR interassay[tiab] OR inter-assay[tiab] OR intraassay[tiab] OR intra-assay[tiab] OR interindividual[tiab] OR inter-individual[tiab] OR intraindividual[tiab] OR intra-individual[tiab] OR interparticipant[tiab] OR inter-participant[tiab] OR intraparticipant[tiab] OR intra-participant[tiab] OR kappa[tiab] OR kappa's[tiab] OR kappas[tiab] OR repeatab*[tiab] OR ((replicab*[tiab] OR repeated[tiab]) AND (measure[tiab] OR measures[tiab] OR findings[tiab] OR result[tiab] OR results[tiab] OR test[tiab] OR tests[tiab])) OR generaliza*[tiab] OR generalisa*[tiab] OR concordance[tiab] OR (intraclass[tiab] AND correlation*[tiab]) OR discriminative[tiab] OR "known group"[tiab] OR factor analysis[tiab] OR factor analyses[tiab] OR dimension*[tiab] OR subscale*[tiab] OR (multitrait[tiab] AND scaling[tiab] AND (analysis[tiab] OR analyses[tiab])) OR item discriminant[tiab] OR interscale correlation*[tiab] OR error[tiab] OR errors[tiab] OR "individual variability"[tiab] OR (variability[tiab] AND (analysis[tiab] OR values[tiab])) OR (uncertainty[tiab] AND (measurement[tiab] OR measuring[tiab])) OR "standard error of measurement"[tiab] OR sensitiv*[tiab] OR responsive*[tiab] OR ((minimal[tiab] OR minimally[tiab] OR clinical[tiab] OR clinically[tiab]) AND (important[tiab] OR significant[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR (small*[tiab] AND (real[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR meaningful change[tiab] OR "ceiling effect"[tiab] OR "floor effect"[tiab] OR "Item response model"[tiab] OR IRT[tiab] OR Rasch[tiab] OR "Differential item functioning"[tiab] OR DIF[tiab] OR "computer adaptive testing"[tiab] OR "item bank"[tiab] OR "cross-cultural equivalence"[tiab]

iv) Exclusion filter

"addresses"[PT] OR "biography"[PT] OR "case reports"[PT] OR "comment"[PT] OR "directory"[PT] OR "editorial"[PT] OR "festschrift"[PT] OR "interview"[PT] OR "lectures"[PT] OR "legal cases"[PT] OR "legislation"[PT] OR "letter"[PT] OR "news"[PT] OR "newspaper article"[PT] OR "patient education handout"[PT] OR "popular works"[PT] OR "congresses"[PT] OR "consensus development conference"[PT] OR

"consensus development conference, nih"[PT] OR "practice guideline"[Publication Type]) NOT ("animals"[MeSH Terms] NOT "humans"[MeSH Terms]

Appendix 2: Quality criteria for rating the results of measurement properties

Measurement property	Rating	Quality criteria
Internal Consistency	+	Factor analyses performed on adequate sample size (7 * # items and >100) AND Cronbach's alpha(s) calculated per dimension AND Cronbach's alpha(s) between 0.70 and 0.95
	?	No factor analysis OR doubtful design or method
	-	Cronbach's alpha(s) ≤ 0.70 or ≥ 0.95, despite adequate design and method
	0	No information found on internal consistency
Reliability	+	ICC or weighted Kappa > 0.70
	?	Doubtful design or method (e.g., time interval not mentioned)
	-	ICC or weighted Kappa < 0.70, despite adequate design and method
	0	No information found on reliability
Measurement error	+	MIC > SDC OR MIC outside the LOA
	?	MIC not defined or doubtful design
	-	MIC < SDC OR MIC equals or inside LOA
	0	No information found on measurement error
Structural validity	+	Factors should explain at least 50% of the variance
	?	Explained variance not mentioned
	-	Factors explain <50% of the variance
	0	No information found on structural validity
Hypothesis testing	+	Specific hypotheses were formulated AND at least 75% of the results are in accordance with these hypotheses
	?	Doubtful design or method (e.g., no hypotheses)
	-	Less than 75% of hypotheses were confirmed, despite adequate design and methods
	0	No information found on hypothesis testing
Cross-cultural validity	+	Original factor structure confirmed or no important DIF found between language versions
	?	Confirmatory factor analysis not applied & DIF not assessed
	-	Original factor structure not confirmed or important DIF found between language versions
	0	No information found on cross-cultural validity
Floor and ceiling effects	+	≤15% of the respondents achieved the highest or lowest possible scores
	?	Doubtful design or method

-	>15% of the respondents achieved the highest or lowest possible scores despite adequate design and methods
0	No information found on interpretation

Adapted from Terwee et al J Clin Epidemiol 2007; 60(1): 34-42. and F. Dobson et al. Osteoarthritis and Cartilage 20 (2012) 1548-1562. Content and criterion validity, responsiveness, & interpretability were not reported on in any included studies; hence have been omitted.

ICC= intraclass correlation coefficient, LOA= limits of agreement, MIC= minimal important change, SDC= smallest detectable change.

Appendix 3: Targeting care: tailoring non-surgical management according to clinical presentation.'

Targeting Care

Tailoring Nonsurgical Management According to Clinical Presentation

Jillian Eyles, BAppSc(Physiotherapy)[a,*],
Barbara R. Lucas, BAppSc(Physiotherapy), MEd, MPH[a],
David J. Hunter, MBBS, MSc, PhD, FRACP[b,c]

KEYWORDS

• Osteoarthritis • Clinical predictors • Treatment response

KEY POINTS

- Numerous studies have explored patient characteristics, including body mass index, psychological factors, muscle strength, tibiofemoral alignment, radiographic changes, and signs of inflammation, as potential predictors of response to nonsurgical interventions for the management of osteoarthritis (OA) of the hip and knee.
- Often the sample sizes used by these studies have been inadequate in yielding sufficient numbers of responders to the interventions to allow for analysis of the potential predictors identified.
- Several well-designed studies have been adequately powered to provide some evidence for clinical characteristics that do or do not predict response to nonsurgical interventions for participants with hip and knee OA.

INTRODUCTION

The activity limitation attributed to osteoarthritis (OA) places it within the world's top 10 most disabling conditions.[1] Globally, OA affects approximately 18.0% of women and 9.6% of men older than 60 years.[2] In 2003 the annual costs to the United States economy of OA and other rheumatic conditions was an estimated $128 billion.[3] These enormous costs are projected to rise steeply with the steadily increasing prevalence of rheumatic conditions.[3]

This prevalent, expensive, disabling disease is incurable, so it follows that current treatments focus on symptomatic relief. Commonly reported treatment goals for this

Disclosures: None.
[a] Department of Physiotherapy, Royal North Shore Hospital, Pacific Highway, St Leonards, New South Wales 2065, Australia; [b] Department of Rheumatology, Royal North Shore Hospital, Pacific Highway, St Leonards, New South Wales 2065, Australia; [c] Sydney Medical School-Northern, Kolling Building Level 7, Royal North Shore Hospital, Reserve Road, St Leonards, New South Wales 2065, Australia
* Corresponding author.
E-mail address: jeyles@nsccahs.health.nsw.gov.au

Rheum Dis Clin N Am 39 (2013) 213–233
http://dx.doi.org/10.1016/j.rdc.2012.11.001 **rheumatic.theclinics.com**
0889-857X/13/$ – see front matter

group include reductions in joint pain, stiffness, activity limitation, participation restriction, and improvements in quality of life and well-being. To assist clinicians in achieving these goals with their patients, numerous international evidence-based guidelines for management of hip and knee OA have become available.[4–10] There is uniformity in most of the recommendations made by the guidelines,[11] and agreement that conservative management of hip and knee OA should combine both nonpharmacologic and pharmacologic treatment modalities.[4–10]

The recommendations made in the guidelines for the management of hip and knee OA are broad. The evidence-based, expert consensus guidelines from the Osteoarthritis Research Society International (OARSI) (2008) include no fewer than 20 recommendations for the nonsurgical management of hip and knee OA[4]; however, the treatments are not arranged systematically to indicate the order of priority in which they should be undertaken. With so many recommended management options tabled, it would be advantageous to know which treatments are most likely to be effective for the individual with hip or knee OA according to clinical presentation.

This review examines the evidence available for the identification of clinical characteristics that predict patient response to nonsurgical treatments for hip and knee OA. The summation of this evidence may assist clinicians to target treatments most likely to benefit patients according to clinical presentation, and identify areas for further research.

BODY MASS INDEX

Obesity is a known risk factor for the development of arthritis[12] and is a strong predictor for long-term progression of the disease.[13] There is evidence that obesity is a risk factor for knee OA, but the relationship between obesity and the risk of developing hip OA is less clear.[14,15] International guidelines nonetheless recommend weight reduction in individuals with hip and knee OA who are overweight or obese.[8,9,11,16] There is strong evidence that weight loss is an effective treatment for knee OA, yet little evidence exists regarding weight loss as an effective treatment for obese patients with hip OA.

It seems reasonable that body mass index (BMI) may be a clinical characteristic that predicts response to weight-loss interventions but, surprisingly, evidence exists that it does not. A post hoc analysis of a randomized controlled trial (RCT) involving 111 overweight veterans with knee OA investigated 9 clinical characteristics as possible predictors of weight change between baseline, 16 weeks, and 32 weeks. The minimum amount of weight loss required to define a treatment responder was not provided. Multiregression analysis revealed that BMI was not predictive of weight loss in response to the interventions for overweight veterans with knee OA.[17] The external validity of this study is limited by confining the recruitment of participants to veterans.

Two studies found that BMI was not predictive of response to a Dutch multimodal, stepped-care model of pain management for hip and knee OA. Snijders and colleagues[18] investigated the efficacy of the Dutch model in a cohort of 183 participants with hip and knee OA. The model combined pharmacologic and nonpharmacologic treatments. Two possible definitions of positive treatment response were described: (1) Outcome Measures in Rheumatoid Arthritis Clinical Trials/Osteoarthritis Research Society International (OMERACT-OARSI) Responder Criteria, and (2) patient-reported numeric rating scale (NRS) for pain of 4 or less. At 12-week reassessment, 86 patients were responders according to definition (1), and 71 fulfilled definition (2). BMI was 1 of 11 potential predictors of response included in analyses, and was not

a significant predictor of response to this program[18]; however, in identifying true predictors of response the study was underpowered. A more recent study used the same Dutch model, focusing specifically on a stepped-care protocol used to progress the use of acetaminophen and nonsteroidal anti-inflammatory drugs (NSAIDs) at standardized intervals according to patient-reported pain levels.[19] The definition of treatment responder was a patient-reported NRS pain 4 or less, and 100 participants met this target. The study was underpowered in analyzing 13 patient characteristics, including BMI, as possible predictors of response. Further research is required to determine whether BMI can predict a positive treatment response in this multimodal stepped-care model of pain management.

Two well-powered studies examined the potential of BMI as a predictor of response to cyclooxygenase-2 (COX-2) inhibitors. Bingham and colleagues[20] pooled the results of 2 similar RCTs comparing the efficacy of etoricoxib and celecoxib with that of placebo. The OMERACT-OARSI Responder Criteria determined that 562 participants were responders to the COX-2 inhibitors following 12 weeks of the intervention. BMI, one of 16 variables analyzed as potential predictors of response, failed to predict a positive treatment response to the COX-2 inhibitors.[20] Similar results were found by Detora and colleagues,[21] who combined the results of 3 6-week RCTs comparing the COX-2 inhibitor rofecoxib with placebo in 1501 patients with hip and knee OA. Responder criteria were not defined. Patient data were analyzed according to subgroups representing 14 baseline characteristics including BMI. Analysis of covariance failed to identify any baseline measures associated with treatment response.[21] To date, good evidence exists that baseline BMI does not predict a response in patients with hip and knee OA treated with COX-2 inhibitors.

A single study explored BMI as a predictor of response to intra-articular corticosteroid injection (CSI) for the management of hip OA. Robinson and colleagues[22] followed 120 patients with hip OA for 12 weeks following CSI. Participants were classified as responders to the CSI at 12 weeks if a reduction in baseline Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain subscale of greater than 15% was achieved; 48 participants met this criterion. Logistic regression determined that BMI, one of 14 variables analyzed, was not a significant predictor of response to hip CSI[22]; however, the study was underpowered in detecting true predictors of treatment response.

Four cohort studies explored possible predictors of response to intra-articular (IA) hyaluronic acid derivatives for hip and knee OA. Short-term efficacy and tolerability of IA Hylan G-F 20 were assessed in 4253 patients with symptomatic knee OA.[23] Responder criteria were not defined, and the primary outcome was pain measured at baseline and 3 weeks on a 4-point Likert scale. At 3 weeks after IA Hylan G-F 20, 88.4% of patients assessed their pain as better or much better. Logistic regression of 7 potential predictors of short-term pain reduction determined that underweight patients were more likely than their obese counterparts to report reduced knee pain. The method of recruitment threatens the validity of this evidence; the investigators invited 840 orthopedic surgeons to report on at least 5 consecutive patients receiving Hylan G-F 20 for relief of knee OA pain, introducing significant selection bias.

Longer-term outcomes of patients with knee OA receiving IA Hylan G-F 20 were explored in 3 cohort studies.[24–26] A retrospective cohort of 155 patients with knee OA was reassessed 7 to 14 months following IA Hylan G-F 20.[24] The definition of responder was not specified. Analysis of 16 possible predictors found that BMI was not a significant predictor of patient satisfaction, although this study was underpowered in identifying the possible predictors, and the retrospective design was prone to significant recall bias. Longer-term outcomes of Hylan G-F 20 were also studied

in a small cohort of 32 patients with mild to moderate knee OA 6 months following IA Hylan G-F 20.[25] Clinical response was defined using the OMERACT-OARSI "high improvement" criterion. Only 15 participants were responders, and 8 variables, including BMI, were investigated as predictors of response, leaving the study underpowered for the detection of significant predictors. BMI was not significantly correlated with patient response. A prospective cohort study examining 84 patients with knee OA for 6 months following knee IA Hylan G-F 20 found that Short-Form-36 (SF-36) health survey scores were significantly improved at 6 months after injection.[26] The responder criteria were not described. Three factors, including the subjects' percentage above ideal body weight, were analyzed for correlations with positive treatment outcomes seen in the SF-36 health survey categories Physical Function, Role-Physical, and Role-Emotional. The subjects' percentage above ideal body weight was not predictive of improvement. The high number of patients lost to follow-up (23%) affected the validity of this study. Evidence for BMI and percentage above ideal body weight as clinical characteristics predictive of longer-term response to knee IA Hylan G-F 20 was inconclusive, owing to low validity and power.

BMI was not a significant predictor of response to IA Hylan G-F 20 in people with hip OA. Migliore and colleagues[27] evaluated 250 patients with hip OA who received IA Hylan G-F 20. Treatment response was defined as a 30% or greater improvement in baseline Lequesne scores or NSAID usage at 6 months, but the number of participants classified as responders was unclear. Ten possible predictors of treatment response were analyzed; BMI was not a significant predictor of response to hip IA Hylan G-F 20.[27] The large number of dropouts (42%) affected the validity of this study.

Patients with a lower BMI may be more likely to experience a reduction in chronic knee pain following treatment with glucosamine sulfate. A prospective correlational study of 39 participants with chronic knee pain followed patients receiving 1.5 g glucosamine sulfate daily for 12 weeks.[28] Participants were not required to have been diagnosed specifically with OA, which affected the external validity of this study. The definition of treatment responder was not described, and 7 patient characteristics were examined as potential predictors of reduction of pain rated on a visual analog scale (VAS). The study was underpowered in determining the effects of 7 potential predictors.

To date, most of the evidence suggests that BMI is not a consistent predictor of response to nonsurgical treatments for people with hip and knee OA. Some evidence exists that BMI is not predictive of response to a weight-loss program in overweight veterans with knee OA.[17] There is good evidence that BMI does not predict response to COX-2 inhibitors for the management of hip and knee OA.[20,21] The evidence is weak that BMI is not predictive of treatment response to either a multimodal stepped-care pain management model,[18,19] hip CSI,[22] hip IA Hylan G-F 20,[27] or glucosamine for chronic knee pain.[28] The evidence for BMI as a predictor of response to knee IA Hylan G-F 20 is weak and conflicting.[23–26] Further research is required to determine whether BMI is a clinical characteristic that can foretell a response to nonsurgical treatments for people with hip and knee OA.

PSYCHOLOGICAL FACTORS

Complex interactions exist between psychological factors and perceived symptoms of OA. Compared with their peers, people with OA report an increased prevalence of depression and depressed mood.[29] The intensity of perceived OA pain has been demonstrated to be predictive of depression severity in this cohort.[29] Poor mental health has been associated with worse overall hip and knee OA pain, and deterioration

in mental health has been found to precede short-term exacerbations of OA pain.[30] Treatment of depression in people with arthritis appears to improve depressive symptoms, reduce OA pain, and improve function and quality of life,[31] and therefore is an important consideration in the management of OA.

Many treatments prescribed for hip and knee OA management, particularly exercise, weight-loss programs, and medications; require active participation from the patient. The compliance with and efficacy of these treatments may be influenced by the individual's mental state as to how rehabilitation outcomes are affected. The prospective cohort study "Predictors for response to rehabilitation in patients with hip or knee OA"[32] featured 250 patients with hip and knee OA who participated in a 3- to 4-week multimodal rehabilitation program combining exercise therapy, hydrotherapy, relaxation strategies, distraction techniques, patient education, manual therapy, thermotherapy, and electrotherapy. Participants were assessed at baseline and 6 months following the program. Three different definitions of treatment responder were used: (1) the minimal clinically important difference (MCID) (18%) improvement shown on the WOMAC, (2) improvement on the Transition scale, and (3) MCID improvement on WOMAC and improvement on the Transition scale. The transition scale was described as a measure of the current state of health of the OA joint compared with its state 6 months earlier.[32] There were 21 personal, lifestyle, and psychological measures investigated as potential predictors of the 3 definitions of responder. Depression and anxiety were evaluated using the Hospital Anxiety and Depression Scale, and mental health was assessed using the mental component of the SF-36. The absence of depressive symptoms was determined to be a strong predictor of all 3 of the responder definitions, suggesting that depression may hinder the achievement of positive treatment outcomes of patients with hip and knee OA following a 3- to 4-week rehabilitation program. This study did not attempt to answer the question as to why patients did not achieve the same results as their nondepressed counterparts, but one may hypothesize that perhaps those patients with depression have more difficulty complying with a comprehensive rehabilitation program. This area is an interesting one for further research.[32]

The presence of depression may affect the ability of overweight people with OA to lose weight. A post hoc analysis of an RCT aimed to identify predictors of positive treatment response resulting from weight-loss interventions for 111 overweight veterans with knee OA.[17] Veterans were randomized into groups receiving 24 weeks of nutritional counseling, a home exercise program, a combination of both, or usual care. There were no differences in weight loss between intervention groups, and 9 variables were investigated as possible predictors of weight change between baseline, 16 weeks, and 32 weeks of the RCT. The amount of weight loss required to indicate successful treatment response was not indicated. Symptoms of depression were evaluated using The Center for Epidemiologic Studies Depression Scale, which measured 20 items to achieve a score out of 60. The presence of depression was indicated by a score of 16 or greater. The absence of depression was the only independent predictor of weight loss at 16 weeks and 32 weeks.[17] This study is limited by its failure to define treatment responders; however, it does suggest that depressive symptoms may limit the ability of veterans to lose weight.

Depression and anxiety did not seem to predict the treatment response of patients with knee OA to CSI. A small study of 59 patients with knee OA receiving CSI examined 10 possible predictors of a favorable response, defined as a 15% or greater reduction of pain rated on VAS, to injection of methylprednisolone acetate.[33] The Hospital Anxiety and Depression score at baseline was not found to consistently predict treatment response. Given that 59 patients were used to investigate

10 predictors of response, this study was underpowered in detecting meaningful effects of the potential predictors.

Mental health scores do not seem to predict response to a combined nonpharmacologic and pharmacologic pain-management program. Predictors for response to analgesics were explored in relation to a cohort study of 347 patients investigating treatment outcomes of a stepped model of care for hip and knee OA. The model initially offered education, lifestyle and weight-loss advice, physiotherapy, and acetaminophen, then progressed to other medications at intervals as guided by a pain NRS.[19] Treatment response was defined as achievement of pain NRS of 4 or greater, and there were 100 responders. Thirteen possible predictors of response were explored, including mental health. The SF-36 questionnaire was used to assess health-related quality of life, and the mental component summary (MCS) scores of the SF-36 were used to reflect mental health. Mental health rated by the MCS was not a significant predictor of response to the stepped model of pharmacologic pain management for patients with hip and knee OA; however, this study was underpowered regarding the analysis of 13 possible predictors.

Self-reported participant mood failed to predict treatment outcome in a small crossover RCT of 11 patients with osteoarthritis receiving 2 different NSAIDs.[34] During 2 treatment periods of 4 weeks' duration, participants received ketoprofen and piroxicam. A 4-week washout period followed the initial drug treatment before commencement of the second drug. Participants were classified as treatment responders if they showed 30% or better improvement of 5 of the 7 variables measured at baseline, including pain, tenderness, swelling, patient and physician global assessments, acute-phase protein levels, and disability. There were 20 baseline variables explored as possible predictors of response including mood, assessed using an 18-item questionnaire. Mood was not a significant predictor of treatment outcome; however, the small sample size of this study leaves it underpowered for the detection of meaningful effects of the predictors investigated.

In summary, 2 well-designed, adequately powered studies used specific measures of depression that were predictive of response to intervention.[17,32] Both studies demonstrated the relationship between the absence of depressive symptoms and positive nonpharmacologic treatment outcomes. The treatments investigated in these studies included a comprehensive rehabilitation program and weight-loss interventions. These treatment modalities require high levels of active participation of the patients involved, which may be affected by the presence of depressive symptoms. Of interest, the 3 studies investigating drug-therapy regimes, perhaps not requiring such a high level of active participation by the subjects, consistently found different measures of psychological factors incapable of predicting treatment response.[19,33,34] Two of these studies were underpowered[33,34] and the third, which was inadequately powered, did not measure depression specifically.

MUSCLE STRENGTH

In view of the biomechanical influence and protective functions of skeletal muscles surrounding joints, muscle weakness is considered to be an important possible factor in the development and progression of OA. Evidence for the significance of muscle strength in the pathogenesis of OA remains unclear.[35,36] Higher quadriceps strength may have a protective effect against the development of symptomatic OA.[36] Whether muscle weakness precedes the onset of OA, or if it is a feature of already established disease seen on radiography, or is only related to the onset of pain and other symptoms, is an area for further research.

The evidence for the role of muscle strength in the progression of OA is varied. Limited evidence exists to support muscle strength as a predictor of progression of knee OA.[13] Nevertheless, over time people with knee OA who have greater quadriceps strength report less pain and superior functional ability compared with their weaker counterparts.[37] Quadriceps strength has been studied widely in relation to knee OA; however, muscles around the hip that stabilize the pelvis also have an effect on adduction forces around the knee, which may result in increased compression of the medial compartment[38] and influence the pathogenesis and progression of OA. Hip OA has also been associated with significantly reduced lower limb muscle strength[39]; however, limited evidence is available to explain the role of hip and thigh musculature in the development and progression of the disease. Further research is required to explain this possible relationship.

Treatments for hip and knee OA have long included specific exercises designed to strengthen muscles surrounding the joints involved. High-level evidence exists regarding the reduction of pain and dysfunction in knee OA through therapeutic exercises.[40] The evidence to date for the efficacy of exercise in hip OA is less convincing,[41] yet exercise is often prescribed. Wright and colleagues[42] published a study aiming to identify baseline characteristics of patients with hip OA likely to respond favorably to physical-therapy interventions. As part of a larger RCT, 91 patients were randomized to groups receiving manual therapy, exercise therapy, a combination of both, or usual care. The OMERACT-OARSI Responder Criteria determined treatment responders. Ten variables were analyzed as predictors of treatment response. Measures of muscle strength using a hand-held dynamometer were not predictive of treatment success. Only 22 of the 68 participants were responders, which left the study underpowered in identifying predictors of response.[42]

There has been recent interest in the nature of lower limb muscle weakness in people with knee OA. Decreased quadriceps strength in knee OA has been attributed to both loss of muscle cross-sectional area[43] and reduced ability to activate the muscles.[44] In a cohort of 111 subjects taken from a larger RCT, baseline ability to activate quadriceps was examined as 1 of 9 possible predictors of changes in strength of the muscle following a 6-week exercise program for subjects with knee OA. Primary outcome measures were quadriceps strength and quadriceps activation, measured using a burst-superimposition maximum isometric quadriceps torque test; however, a definition of treatment response was not identified. Although activation of lower quadriceps was associated with lower strength, baseline activation of quadriceps did not predict the magnitude of gain in quadriceps strength following exercise therapy.[45] These results suggest that patients with OA should benefit from strengthening exercises regardless of baseline activation of quadriceps.

Baseline muscle strength does not seem to predict the degree of symptomatic relief achieved following a weight-loss program in obese people with knee OA. The 192 participants, who were part of a larger RCT, were randomized to 2 different dietary interventions. Significant response to the interventions included the OMERACT-OARSI Responder Criteria and improvement on Knee Injury and Osteoarthritis Outcome (KOOS) scores. Although weight loss was achieved in most of the subjects, only 64% achieved the OMERACT-OARSI responder criterion. There were 23 variables investigated as possible predictors of response to the weight-loss programs, including measurements of baseline hamstrings and quadriceps strength using isometric dynamometry. Baseline muscle strength was not predictive of symptomatic relief in response to the weight-loss program.[46] The study was underpowered in detecting significant predictors from a possible 23 variables.

One study investigated muscle strength as a predictor of response to a pharmacologic agent. Jones and Doherty[33] performed a crossover RCT comparing CSI with saline (placebo) in 59 subjects with knee OA. Ten possible predictors of treatment response were analyzed. Treatment response, defined as a decrease of 15% or more in pain rated on a VAS, was not predicted by baseline quadriceps strength measured using a commercial strain gauge. This study was significantly underpowered for the analysis of 10 predictors of response.

Further research into the role of muscle strength in the pathogenesis of OA and subsequent progression of the disease may be helpful in refining recommendations for therapies aimed at OA prevention and further joint deterioration as a consequence of OA. To date, muscle strength has not been demonstrated to predict response to nonsurgical interventions for hip and knee OA.

TIBIOFEMORAL JOINT ALIGNMENT

Varus (bow-legged) or valgus (knock-kneed) tibiofemoral joint alignments are clinical characteristics observed in some people with knee OA. Joint alignment affects the distribution of load borne by the medial and lateral compartments of the articular surface of the knee. Static knee alignment is conventionally determined using full-length weight-bearing radiographs of the lower limb with knees extended. Lines are drawn from the center of the femoral head to the talus through the middle of the femoral and tibial shafts to indicate the load-bearing mechanical axis, then measurements are made of various angles subtended from where those lines intersect.[47–49] Neutral alignment is commonly defined as 0° to 2° of varus,[50] meaning that in a normal knee the mechanical axis passes medial to the knee joint, resulting in 60% to 70% of weight-bearing forces passing through the medial articular surface.[51] Varus malalignment results in higher loads borne through the medial compartment of the knee, whereas increased compressive forces through the lateral articular surface accompany valgus malalignment.

Dynamic knee alignment can be assessed using 3-dimensional gait analysis. In varus knees the measurement of knee-adduction moment during the stance phase of walking is an indirect measure of joint compressive forces sustained within the medial tibiofemoral joint compartment.[49,52,53] Static and dynamic alignment is an important consideration, given that altered distribution of forces placed through the joint surface may lead to damage of articular structures, possibly increasing the risk for development of OA or worsening existing disease.

It remains unclear as to whether knee-joint malalignment precedes incident knee OA[47,48]; however, varus alignment is considered to be a significant predictor of knee OA disease progression.[13] Knee malalignment has been demonstrated to interact with other risk factors for OA progression, increasing the likelihood of disease acceleration. Possible interactive factors include greater quadriceps strength,[54] the stage of disease observed in the individual,[55] and obesity.[48]

The evidence for the relationship between knee malalignment and reported OA symptoms remains unclear.[55,56] Nevertheless, some nonsurgical treatments in OA management guidelines aim to reduce pain and dysfunction associated with tibiofemoral malalignment. Orthotic bracing, shoe wedges, and muscle strengthening are recommended with a view to improving the biomechanics of the joint.[4,6,7,9,10] Several studies have investigated knee-joint alignment as a predictor response to nonsurgical management of OA. An RCT by Lim and colleagues[57] examined the effect of a 12-week quadriceps-strengthening program on knee-adduction moment, pain, and function in 107 subjects with knee OA. Knee alignment was assessed on

radiographs, and participants were stratified according to whether they had more neutral (<5°) or more varus (≥5°) alignment. Specific responder criteria were not described. Patients in the strengthening group achieved significant improvements in strength regardless of alignment. Self-reported function, performance measures, and knee-adduction moment determined using 3-dimensional gait analysis were unchanged by the intervention in both alignment groups. Pain, assessed using the WOMAC pain subscale, was significantly improved in the strengthening group subset that was more neutrally aligned. Neutral knee-joint alignment may mediate improvements in knee OA pain following a 12-week quadriceps-strengthening program.

Immediate changes in static alignment and knee-adduction moment were not predictive of response to lateral-wedge insoles at 3 months. A cohort of 40 volunteers with knee OA were provided with laterally wedged insoles to assess the immediate effects of the insoles on knee OA pain, knee-adduction moment, and static alignment.[49] The lateral wedges immediately reduced knee-adduction moment calculated using 3-dimensional gait analysis and walking pain measured using the WOMAC pain subscale, but had no effect on static alignment as determined on full-length leg radiographs. Alignment was defined as the angle subtended by the intersection of the femoral and tibial mechanical axes. Varus malalignment was determined when the angle was less than 180°, with valgus indicated by an angle of greater than 180°. After 3 months of wearing the insoles, significant improvements in pain and function persisted. A definition of treatment responder was not specified; nevertheless, 10 predictors at baseline of outcome after 3 months were explored. Neither immediate changes in static alignment nor knee-adduction moment were predictive of decreased pain and improvement in function 3 months following the intervention.[49] The size of this cohort limited the ability of the study to identify true predictors of response to the intervention.

A larger RCT of 192 obese subjects with knee OA allocated patients to 2 different weight-loss interventions.[46] Knee-joint alignment was assessed using a "Plug-in-Gait" model with a 6-camera stereophotogrammetric system and markers on anatomic landmarks. A knee was categorized as varus when alignment was greater than 0°, and valgus if less than 0°. Baseline knee alignment was one of 23 variables examined as possible predictors; however, it failed to predict improvements in KOOS or achievement of OMERACT-OARSI Responder Criteria following weight-loss interventions.[46] In view of the fact that only 64% of patients were treatment responders according to OMERACT-OARSI criteria, the study was underpowered in detecting effects of significant predictors.

It is interesting to consider the definitions of knee malalignment used in the 3 studies discussed above. Lim and colleagues[57] used a more extreme definition of 5° or more to indicate varus malalignment. By contrast, the 2 other studies categorized subjects to knee-malalignment groups if the mechanical axis did not appear as a straight line.[46,49] This disparity may have increased the severity of malalignment observed within the participants assigned to the varus group investigated by Lim and colleagues, compared with the subjects categorized to knee-malalignment groups in the other studies. Participants with varus malalignment studied by Lim and colleagues[57] did not experience improvements in pain following strength training, whereas neutrally aligned subjects reported significant pain reduction. Perhaps the higher severity of varus malalignment was key to the determination of knee-joint alignment as a predictor of outcome after intervention in this study. Future research considering knee malalignment as a predictor of treatment response to conservative treatments should consider carefully the definition of joint alignment.

ASSESSMENT BY RADIOGRAPHY AND MAGNETIC RESONANCE IMAGING

The presence of radiographic osteophytes (OP) and joint-space narrowing are commonly used to diagnose OA. These features are combined to determine radiographic disease severity according to scoring systems such as the Kellgren-Lawrence grade (KLG).[58] Despite known limitations, radiographs are inexpensive, accessible, and easy to interpret, so are commonly used in research for the classification of subjects to determine eligibility, and for stratification of samples according to radiographic severity. Radiographic joint-space width (JSW) or minimum joint-space width (mJSW) is recommended for use in clinical trials; however, magnetic resonance imaging (MRI) is preferred particularly for the assessment of cartilage morphology.[59]

Relatively few articles analyze radiographic severity of hip and knee OA as possible predictors of response to nonsurgical, nonpharmacologic treatments. Two of the 3 articles doing so examined the ability of radiologic and MRI OA severity to predict response to weight-loss interventions. A small RCT of 30 obese female participants with knee OA compared 2 dietary weight-loss interventions.[60] Within the intervention group 90% of participants achieved a clinically significant weight reduction of greater than 10%, and 33% had a 50% improvement in symptoms of knee OA. A strict definition of treatment responder was not provided. Structural joint damage was assessed at baseline using both the KLG classification and low-field MRI (0.2 T) to assess various measures of cartilage abnormalities, bone marrow lesions, effusions, and synovitis of the medial, lateral, and patellofemoral compartments of the knee. Five baseline radiographic characteristics and clinical outcomes following the weight-loss interventions were investigated for correlations. None of the imaging variables were able to forecast symptomatic response to treatment[60]; however, this study was likely underpowered in identifying significant predictors.

A second RCT randomized 192 obese patients with knee OA into 8 weeks of 2 experimental dietary interventions.[46] Results were calculated for the entire cohort, as the method of weight loss was not relevant for this analysis. OA symptoms were evaluated at baseline and at 16 weeks using the OMERACT-OARSI Responder Criteria and changes in KOOS. High-field MRI was assessed using the Boston-Leeds Osteoarthritis of the Knee Score (BLOKS) to measure joint damage at baseline. Conventional radiography determined the baseline KLG and mJSW. MRI and radiographic measures failed to find any relationship between variables assessing structural damage to the knee and symptomatic improvements following the dietary interventions.[46] Only 64% of patients were treatment responders according to OMERACT-OARSI criteria, therefore this study may also be insufficiently powered to detect the effects of 23 potential predictors.

A third study examining the ability of radiographic features to predict response to nonsurgical, nonpharmacologic interventions was conducted by Hinman and colleagues.[49] A cohort of 40 patients with knee OA wore full-length 5° lateral-wedge insoles for 3 months. Improvements were observed in WOMAC pain and function subscales following the intervention. Tibiofemoral OA severity was assessed at baseline using the KLG scoring system. Following analysis of 10 possible predictors of outcome, greater disease severity indicated by higher KLG scores was predictive of worse pain at 3 months. This study does not define responder criteria, and the small sample size reduced the ability to identify predictors of response to lateral-wedge insoles.

Two studies examined radiographic severity using KLG as potentially predictive of response to interventions combining both nonpharmacologic and drug therapies for hip and knee OA. Both investigated cohorts of patients with hip and knee OA

participating in a Dutch multimodal, stepped-care pain-management program.[18,19] During the 12-week program subjects received standardized nonpharmacologic management and pain-relieving medications, prescribed and altered at set intervals depending on self-reported pain at reassessment. The definition of positive treatment response in the initial cohort of 183 patients was fulfillment of either the OMERACT-OARSI Responder Criteria or NRS of 4 or less.[18] The later study of 347 subjects required NRS of 4 or less at 12 weeks to indicate successful response to the intervention.[19] Both studies analyzed OA severity as determined by KLG scores as possible predictors of positive treatment outcomes. The first study tested 11 possible predictors of response to intervention, and found that disease severity did not forecast improvements in overall pain and function as a result of the 12-week pain management program.[18] In the second study, 13 predictors were tested for correlation with treatment response at the 4 different steps of the treatment model. Greater OA severity was independently associated with a higher chance of pain relief achieved in response to the use of acetaminophen.[19] This correlation was discovered because unlike in the first study, the predictors of response were tested at each of the separate steps of the program. There were 59 responders to acetaminophen, so the study was underpowered for the testing of 13 predictors. Although the evidence is tenuous, this finding lends support to the recommendations made by international OA management guidelines to trial acetaminophen as a first-line pharmacologic treatment of hip and knee OA,[4–10] even in those patients with severe disease.

Evidence to the contrary was presented by Case and colleagues[61] in the results from a double-blind, placebo-controlled RCT comparing the efficacy of acetaminophen and diclofenac sodium for the management of pain in knee OA. Eighty-two patients were randomized to 3 groups receiving either one of the drugs or placebo. The primary outcome at baseline, 2 weeks, and 12 weeks was the WOMAC scale. The diclofenac sodium group alone achieved significant improvement ($\geq$20%) in all 3 WOMAC subscales following the intervention. The subjects were stratified according to prestudy medication, baseline pain, and disease severity indicated by KLG to identify subsets of patients who were consistent in their response to the treatments. None of the subgroups consistently demonstrated preferential response to acetaminophen or diclofenac sodium. This study suffered from a high number of dropouts (>25%). Three of the 5 subjects who withdrew from the diclofenac sodium treatment arm (n = 25) did so as a result of adverse effects. Despite the evidence presented in this study for the superior efficacy of diclofenac sodium, the relatively high risk of unwanted side effects lends further weight to the OA treatment guidelines that recommend a trial of acetaminophen before commencing NSAID therapy,[4–10] and it can be presumed that this follows regardless of radiographic severity.

Four articles investigated radiologic predictors of response to CSI for hip OA. Of these, only one reported that radiographically determined disease severity was a significant predictor of positive response to steroid injection.[62] This retrospective cohort study reviewed radiographs, radiology reports, and medical records of 361 patients who had received fluoroscopically guided IA methylprednisolone acetate 80 mg, or methylprednisolone with bupivacaine. The definition of treatment responder was a 50% decrease in pain reported on a VAS. An immediate positive response to injection was evident in 68.2% of hips and a delayed response was apparent in 71.4%. OA severity was measured at baseline using KLG classification, and the grades were split into groups for analysis. Multivariate regression determined that radiographic severity of OA was an independent predictor of treatment response. Patients with advanced disease were much more likely to experience both immediate and delayed onset of pain relief. The investigators suggested that people with

advanced hip OA are likely to achieve a better response to CSI than those with mild or moderate disease.[62] Although a sizable number of participants was recruited, these inferences should be considered cautiously in view of the inherent risk of bias associated with the retrospective cohort design of this study.

By contrast, Robinson and colleagues,[22] using a similar fluoroscopically guided injection of methylprednisolone and bupivacaine into the hip joint of 120 people with hip OA, concluded that radiographically determined OA severity was not predictive of response to intervention. This cohort study assessed symptomatic response to 40-mg and 80-mg doses of the steroid. A decrease in the WOMAC pain by greater than 15% was considered to indicate positive treatment response, and 75 patients were classified as responders at 6 weeks. The investigators concluded that the higher dose (80 mg) of methylprednisolone was more effective and lasted longer. Twelve possible predictors of treatment response included KLG scoring. Forward logistic regression found that KLG was incapable of predicting reduced pain in response to hip CSI.[22] This study was underpowered in detecting predictors of response among 12 variables.

Similar conclusions were made regarding a small prospective cohort of 27 patients with hip OA assessed at baseline and 2, 12, and 26 weeks following hip IA lignocaine and methylprednisolone.[63] The main outcome measure was pain measured on VAS. The degree of radiologic severity according to KLG classification and mJSW had no significant bearing on the reported pain relief following steroid injection to the hip; however, the small sample size decreased the power to detect significant predictors. The fourth RCT compared ultrasound-guided CSI with IA hyaluronic acid, saline (control), and standard care (no injection) in 77 subjects with hip OA.[64] Response to treatment was delineated by the OMERACT-OARSI Responder Criteria, and there were 14 responders to steroid injection. CSI was significantly more effective than the 3 other treatments. Univariate regression analysis determined that of 5 predictors analyzed, radiographic severity using Croft grading and mJSW were not predictive of treatment response to CSI; however, the study was underpowered in analyzing 5 predictors.[64] Further research is required to explore the value of radiographic and MRI clinical characteristics indicating disease severity as potential predictors of response to hip CSI.

Three cohort studies attempted to identify radiographic characteristics of patients with hip and knee OA that were predictive of treatment response to IA Hylan G-F 20.[24,25,27] Migliore and colleagues[27] followed 250 patients who received ultrasound-guided IA Hylan G-F 20 into OA hips. Treatment response was defined as improvement of greater than 30% or more in Lequesne index or NSAID use. Significant improvements were reported for all outcome measures at 3, 6, 9, and 12 months when compared with baseline. Multiregression analysis of 8 baseline variables determined that KLG was unable to predict treatment response.[27] A high number of dropouts limited the validity of this study. The second study followed a small cohort of 32 patients with mild to moderate knee OA for 6 months following IA Hylan G-F 20.[25] The OMERACT-OARSI "High Improvement" responder criteria for OA were used to define responders to treatment. Fifteen participants met the responder criteria. Eight predictors of treatment response were explored, including mJSW, which was not predictive of positive response to Hylan G-F 20 injection.[25] The study was underpowered and limited by the exclusion of patients with severe OA. Conrozier and colleagues[24] studied a cohort of 155 patients across the spectrum of mild through severe knee OA. Knee joint-space loss in a single compartment seen on radiographs and meniscal calcinosis noted on MRI scans were predictive of a good outcome after IA Hylan GF-20. The definition of treatment responder was not adequately described in this

retrospective cohort, so it is difficult to define the improvement actually predicted by these measures.[24] Nevertheless, weak evidence exists that knee joint-space loss in a single compartment and meniscal calcinosis may predict response to IA Hylan G-F 20, and this warrants further research.

Bennett and colleagues[28] investigated the symptomatic response of 39 subjects with chronic knee pain treated with 1.5 g oral glucosamine sulfate for 12 weeks. Primary outcome measures at baseline and 3 months included pain VAS rated on movement, VAS for restriction in function, and patient-rated global change score. These outcomes were all found to be significantly improved at 12 weeks, but the responder criteria were not specified. Seven possible predictors of reduced pain and improved functional ability were analyzed using regression modeling. The investigators concluded that lower levels of osteophytes in the patellofemoral joint, BMI, and functional self-efficacy were predictors of successful glucosamine treatment. The presence of osteophytes within the medial and lateral compartments of the tibiofemoral joint was not correlated with response to the intervention.[28] The study was underpowered for the identification of true predictors of response, and the participants did not require formal diagnosis of OA, so these results must be viewed accordingly.

Overall, there was weak evidence that radiographic measures of OA severity may have predictive value in the identification of potential responders to lateral-wedge insoles,[49] CSI,[62] and glucosamine sulfate. MRI assessment was predictive of response in a single study concerned with Hylan G-F 20 injections in the knee.[24] A greater number of studies exist that were unable to predict response to treatment based on radiographic disease severity or MRI. There is good evidence that KLG scores are not predictive of response to hip CSI.[22] Further research is required to clarify the roles performed by radiography and MRI regarding clinical characteristics that predict response to the nonsurgical treatment of hip and knee OA.

INFLAMMATION

Abnormal progressive remodeling of joint tissues occurs in response to local inflammatory processes arising within osteoarthritic joints.[65] Physical examination may reveal clinical signs such as presence of joint swelling, effusion, and heat. With recent improvements in imaging techniques, synovial hypertrophy has become a surrogate marker of local inflammation within a joint.

Signs of inflammation were examined as potential predictors of response to weight-loss interventions for participants with knee OA. The clinical cohort described by Gudbergsen and colleagues[46] participated in a 4-month dietary intervention. Responders were required to fulfill the OMERACT-OARSI Responder Criteria. Joint-damage severity was assessed on MRI using the BLOKS, which included scoring for synovitis and effusion. Although synovitis and effusion were not predictive of OMERACT-OARSI response, there was some evidence that the effusion score correlated with changes in the KOOS activities of daily living score from baseline to 4 months. Responder criteria for the KOOS score were not provided. There were 23 variables assessed as potential predictors of response; 123 patients were responders, therefore this study was insufficiently powered for this many predictors. The presence of inflammatory markers such as effusion and synovitis requires further investigation as predictors of symptomatic response to weight-loss interventions for overweight patients with OA.

Systemic pharmacologic agents such as NSAIDs and COX-2 inhibitors are prescribed for their analgesic properties and also to reduce inflammatory activity in affected joints. Two studies found that the presence of swelling was not predictive

of response to these drug therapies. The data of 3 6-week RCTs comparing rofecoxib with placebo were combined to analyze the consistency of response of patients with hip or knee OA classified into subgroups determined by 14 demographic and disease factors.[21] Three outcome measures were analyzed in relation to the subgroups: pain walking on a flat surface (WOMAC), patient global assessment of response to therapy, and global assessment of disease status. The definition of treatment responder was not provided. Overall, the subgroups did not show consistent interactive effects with all 3 outcome measures. The absence of knee swelling in participants with knee OA significantly correlated with improved scores on the patient global assessment of response to therapy, but not the 2 remaining outcome measures.[21] Another study investigated swelling among numerous possible predictors of response of patients with OA and rheumatoid arthritis to ketoprofen and piroxicam.[34] The trial was very small with only 11 participants with OA, so was underpowered in determining significant predictors of response.[34] Further investigation into signs of inflammation as possible predictors of response to NSAIDs and COX-2 inhibitors would be helpful to the clinician attempting to tailor pharmacologic management according to clinical presentation.

IA corticosteroids aim to directly reduce inflammatory processes occurring within joint tissues. An RCT by Chao and colleagues[66] examined inflammatory characteristics assessed on ultrasonography as predictors of response to IA corticosteroid injection for knee OA. Participants were categorized as inflammatory if synovial hypertrophy (synovitis) with or without effusion was detected on gray-scale ultrasound examination of the affected knee(s) at baseline. Within the intervention group, 16 patients presented with synovitis on ultrasonography and 18 did not. At 4 weeks there were no significant differences between the inflammatory and noninflammatory subgroups. Significantly lower WOMAC pain scores of the noninflammatory subgroup at 12 weeks suggested that those without inflammatory characteristics experienced prolonged beneficial effects from corticosteroids. The presence of effusion had no influence on response to corticosteroid injection.[66]

The presence of hip-joint synovitis on ultrasound assessment of patients with hip OA was predictive of treatment response to CSI. An RCT compared standard care (no injection), injection of normal saline (placebo), nonanimal stabilized hyaluronic acid, and methylprednisolone acetate.[64] Of the participants receiving CSI, 14 participants were classified as responders according to the OMERACT-OARSI criteria. The investigators concluded that synovitis was predictive of response at 4 and 8 weeks following injection; however, this study was underpowered in establishing clear associations between the 5 variables analyzed as possible predictors. By contrast, Robinson and colleagues[22] found that evidence of hip synovitis and effusion on ultrasonography were not predictive of clinical response to IA methylprednisolone injection. The cohort study defined response to intervention as greater than 15% reduction in baseline WOMAC pain score at 6 and 12 weeks following injection. This study was also underpowered in identifying predictors of response. Further research using greater numbers of subjects is required to explore ultrasound-determined inflammatory characteristics as predictors of response to CSI of osteoarthritic hips.

Inflammatory characteristics identified on physical examination of patients with knee OA failed to predict response to CSI. The presence of local inflammation indicated by knee-joint fluid, local heat, synovial thickening, and stiffness were explored as possible predictors of response to IA methylprednisolone in an RCT of 59 participants with symptomatic knee OA.[33] No predictors of response were identified, perhaps as a result of this study being underpowered. Pendleton and colleagues[67] examined similar clinical signs of inflammation, namely presence of heat, effusion,

and synovial thickening, in addition to the presence of effusion and synovitis on knee ultrasonography, as predictors of improvements in baseline WOMAC pain scale 1 and 6 weeks following CSI. The presence of heat was associated with a 29% greater reduction in night pain; otherwise, clinical and ultrasonographic inflammatory signs were not predictive of response. The study was underpowered and did not publish any measures of data variability. Adequately powered, well-designed research is necessary to determine whether clinical and ultrasonographic signs of inflammation are predictive of outcomes following CSI for knee OA.

Moderate effusion was associated with a good outcome following IA injection with Hylan G-F 20 in patients with symptomatic OA. Conrozier and colleagues[24] followed a cohort of 155 patients who received 3 IA Hylan G-F 20 injections and were evaluated 7 to 14 months later. Treatment outcomes included patient satisfaction, safety, and changes in pain and function, which were assessed on 4-point Likert scales. This study was limited by the lack of validated outcome measures and the retrospective study design.

Only one study investigated signs of inflammation as predictive of outcome to nonsurgical, nonpharmacologic intervention. There is weak evidence that that synovitis and effusion seen on MRI are unable to predict response to weight loss in patients with OA.[46] Numerous studies were concerned with signs of inflammation as predictors of outcomes to pharmacologic agents, but few were sufficiently powered. Some evidence exists that knee-joint swelling may predict good outcomes from rofecoxib[21] and that patients without synovitis observed on ultrasonography experience prolonged pain relief following CSI injection in comparison with patients with knee OA presenting with synovitis.[66] The evidence for synovitis on ultrasonography as a predictor of response for outcomes following hip CSI is conflicting.[22,64] There is little evidence to support the use of clinical inflammatory signs as predictors of response to CSI for knee OA.[33,67] Further research is required to determine whether signs of inflammation are useful predictors of response to conservative therapies for people with hip and knee OA.

OTHER CLINICAL CHARACTERISTICS THAT MAY PREDICT RESPONSE TO INTERVENTION

For the purposes of this review, the authors have selected patient characteristics deemed worthy of examination as potential predictors of response to interventions for those with hip and knee OA. There is a wider range of presenting features than those covered here, analyzed as potential predictors of response and further discussed in the literature. Among the articles identified through literature searches for the predictors chosen, age and gender were commonly analyzed as potential predictors of response to intervention, but appeared to hold little predictive capacity overall. Four well-powered studies investigating predictors of response to COX-2 inhibitors,[20,21] a rehabilitation program,[32] and exercise therapy[45] provided moderate to good evidence that age was not a powerful predictor of response to these interventions. Further investigation of age as a predictor of response to alternative interventions for patients with hip and knee OA is justified.

One adequately powered study determined female gender to be a characteristic predictive of treatment success following participation in a rehabilitation program.[32] By contrast, 3 well-powered studies found that gender was not predictive of treatment success for COX-2 inhibitors and exercise therapy.[20,21,45] Additional research into gender as a predictor of treatment response to different nonsurgical modalities is required.

WOMAC subscales of pain and function are often used as primary outcome measures in OA research. Of the studies extracted from literature searches performed

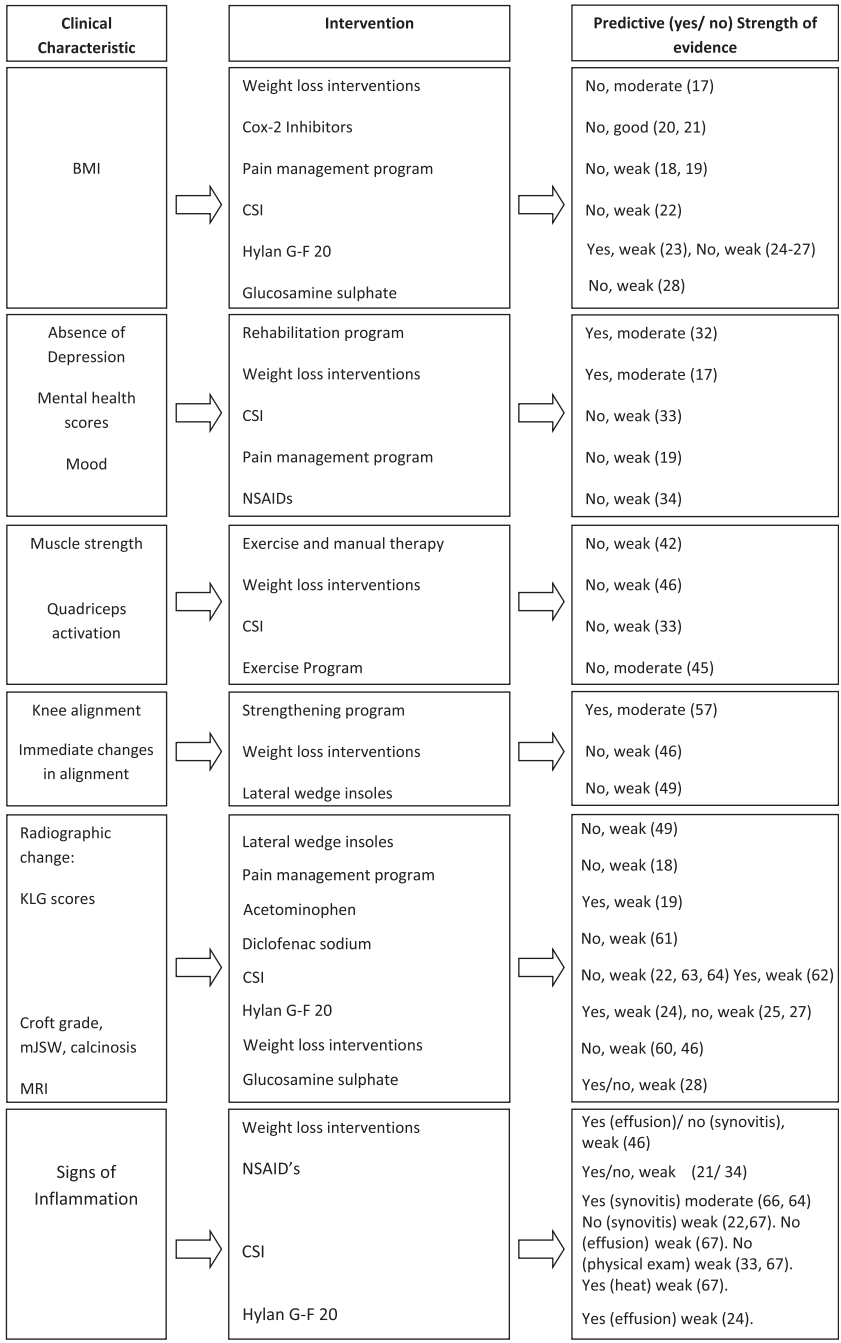


Fig. 1. Summary of evidence available for particular features of clinical presentation shown by people with hip and knee osteoarthritis. BMI, body mass index; CSI, intra-articular corticosteroid injection; KLG, Kellgren-Lawrence Grade; mJSW, minimum joint space width; MRI, magnetic resonance imaging; NSAIDs, nonsteroidal anti-inflammatory drugs.

around the chosen predictors, 2 well-designed studies examined WOMAC pain and function scores as predictors of response to COX-2 inhibitors. One found that that baseline WOMAC pain was not predictive of response to etoricoxib and celecoxib. Lower levels of function on the WOMAC decreased the odds of response to the drugs, but the difference in WOMAC function scores between responders and nonresponders was not clinically significant.[20] The second study concluded that baseline WOMAC function was not predictive of response to rofecoxib.[21] Although baseline WOMAC pain and function scores were not predictive of response to COX-2 inhibitors, these measures may prove to be interesting predictors of response to different nonsurgical interventions in other research.

SUMMARY

This review identified and summarized the evidence available for particular features of clinical presentation shown by individuals with hip and knee OA that were predictive of response to nonsurgical interventions. The studies are summarized in **Fig. 1**. Good evidence exists that BMI is not predictive of response to COX-2 inhibitors for hip and knee OA,[20] and there is moderate evidence that BMI does not predict weight reduction following weight-loss interventions for overweight people with knee OA.[17] There is some evidence to suggest that the absence of depressive symptoms predicts successful outcomes from both weight-loss interventions in overweight people with knee OA[17] and a 3- to 4-week rehabilitation program for participants with hip and knee OA.[32] Moderate evidence was cited that activation of quadriceps muscle was not predictive of improvements in quadriceps strength attained by participants with knee OA during a strengthening program.[45] Patients with medial knee OA who were neutrally aligned were more likely than their more varus-aligned counterparts to achieve significant pain relief following a quadriceps-strengthening program.[57] Evidence was lacking for any radiographic or MRI changes that were significant predictors of response to nonsurgical interventions; however, patients with knee OA presenting without inflammatory characteristics on ultrasonography (synovitis) were more likely to experience prolonged benefit from CSI than were inflammatory patients.[66]

The practice of analyzing patient characteristics as potential predictors of response to interventions is becoming increasingly popular. Researchers attempting to identify predictors of clinical response to nonsurgical treatments for hip and knee OA require the use of larger sample sizes, or restriction of the number of variables analyzed such that 10 to 15 responders are studied per possible predictor.[68] Identification of further characteristics capable of predicting response to intervention would indeed provide clinicians with additional tools to tailor the nonsurgical care of patients with hip and knee OA according to their clinical presentation.

REFERENCES

1. World Health Organization. Chronic diseases and health promotion: chronic rheumatic conditions. Available at: http://www.who.int/chp/topics/rheumatic/en/. Accessed September 17, 2012.
2. Murray C, Lopez A. The global burden of disease. A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020. Cambridge (MA): Harvard School of Public Health on behalf of the World Health Organization and The World Bank; 1996.
3. Centers for Disease Control and Prevention (CDC). National and state medical expenditures and lost earnings attributable to arthritis and other rheumatic conditions—United States. MMWR Morb Mortal Wkly Rep 2007;56:4–7.

4. Zhang W, Moskowitz R, Nuki G, et al. OARSI recommendations for the management of hip and knee osteoarthritis, Part II: OARSI evidence-based, expert consensus guidelines. Osteoarthritis Cartilage 2008;16(12):137–62.
5. Zhang W, Doherty M, Arden NK, et al. EULAR evidence based recommendations for the management of hip osteoarthritis: report of a task force of the EULAR Standing Committee for International Clinical Studies Including Therapeutics (ESCISIT). Ann Rheum Dis 2005;64:669–81.
6. Pendleton A, Arden N, Dougados M, et al. EULAR recommendations for the management of knee osteoarthritis: report of a task force of the Standing Committee for International Clinical Studies Including Therapeutic Trials (ESCISIT). Ann Rheum Dis 2000;59:936–44.
7. Royal Australian College of General Practitioners. Guideline for the non-surgical management of hip and knee osteoarthritis. South Melbourne (Australia): Royal Australian College of General Practitioners; 2009.
8. Zhang W, Nuki G, Moskowitz RW, et al. OARSI recommendations for the management of hip and knee osteoarthritis: part III: changes in evidence following systematic cumulative update of research published through January 2009. Osteoarthritis Cartilage 2010;18(4):476–99.
9. NICE and Royal College of Physicians Guidelines on Osteoarthritis. Osteoarthritis—national clinical guideline for care and management in adults. United Kingdom 2008. Available at: http://guidance.nice.org.uk/CG59. Accessed September 18, 2012.
10. Hochberg MC, Altman RD, April KT, et al. American College of Rheumatology 2012 recommendations for the use of nonpharmacologic and pharmacologic therapies in osteoarthritis of the hand, hip, and knee. Arthritis Care Res (Hoboken) 2012;64(4):465–74.
11. Misso M, Pitt V, Jones K, et al. Quality and consistency of clinical practice guidelines for diagnosis and management of osteoarthritis of the hip and knee: a descriptive overview of published guidelines. Med J Aust 2008; 189(7):394–9.
12. Janssen I, Mark AE. Separate and combined influence of body mass index and waist circumference on arthritis and knee osteoarthritis. Int J Obes 2006;30(8): 1223–8.
13. Chapple CM, Nicholson H, Baxter GD, et al. Patient characteristics that predict progression of knee osteoarthritis: a systematic review of prognostic studies. Arthritis Care Res (Hoboken) 2011;63(8):1115–25.
14. Heliövaara M, Mäkelä M, Impivaara O, et al. Association of overweight, trauma and workload with coxarthrosis: a health survey of 7,217 persons. Acta Orthop Scand 1993;64(5):513–8.
15. Karlson E, Mandl L, Aweh G, et al. Total hip replacement due to osteoarthritis the importance of age, obesity, and other modifiable risk factors. Am J Med 2003; 114(2):93–8.
16. Brand C, Hunter DJ, Hinman RS, et al. Improving are for people with OA of the hip and knee: how has national policy for OA been translated into service models in Australia. Int J Rheum Dis 2011;14:181–90.
17. Wolf S, Foley S, Budiman-Mak E, et al. Predictors of weight loss in overweight veterans with knee osteoarthritis who participated in a clinical trial. J Rehabil Res Dev 2010;47(3):171–81.
18. Snijders GF, den Broeder AA, van Riel PL, et al. Evidence-based tailored conservative treatment of knee and hip osteoarthritis: between knowing and doing. Scand J Rheumatol 2011;40(3):225–31.

19. Snijders GF. Treatment outcomes of a Numeric Rating Scale (NRS)-guided pharmacological pain management strategy in symptomatic knee and hip osteoarthritis in daily clinical practice. Clin Exp Rheumatol 2012;30(2):164–70.
20. Bingham CO 3rd, Smugar SS, Wang H, et al. Predictors of response to cyclooxygenase-2 inhibitors in osteoarthritis: pooled results from two identical trials comparing etoricoxib, celecoxib, and placebo. Pain Med 2011;12(3):352–61.
21. Detora L, Krupta D, Bolognese J, et al. Rofecoxib shows consistent efficacy in OA clinical trials, regardless of specific patient demographic and disease factors. J Rheumatol 2001;28(11):2494–503.
22. Robinson P, Keenan AM, Conaghan PG. Clinical effectiveness and dose response of image-guided intra-articular corticosteroid injection for hip osteoarthritis. Rheumatology 2007;46(2):285–91.
23. Kemper F, Gebhardt U, Meng T, et al. Tolerability and short-term effectiveness of Hylan G-F 20 in 4253 patients with osteoarthritis of the knee in clinical practice. Curr Med Res Opin 2005;21(8):1261–9.
24. Conrozier T, Mathieu P, Schott A, et al. Factors predicting long-term efficacy of Hylan GF-20 viscosupplementation in knee osteoarthritis. Joint Bone Spine 2003;70:128–33.
25. Anandacoomarasamy A, Bagga H, Ding C, et al. Predictors of clinical response to intraarticular Hylan injections—a prospective study using synovial fluid measures, clinical outcomes and magnetic resonance imaging. J Rheumatol 2008;35(4):685–90.
26. Goorman S, Watanabe T, Miller E, et al. Functional outcome in knee osteoarthritis after treatment with Hylan G-F 20: a prospective study. Arch Phys Med Rehabil 2000;81:479–83.
27. Migliore A, Tormenta S, Massafra U, et al. Intra-articular administration of Hylan G-F 20 in patients with symptomatic hip osteoarthritis: tolerability and effectiveness in a large cohort study in clinical practice. Curr Med Res Opin 2008; 24(5):1309–16.
28. Bennett AN, Crossley KM, Brukner PD, et al. Predictors of symptomatic response to glucosamine in knee osteoarthritis: an exploratory study. Br J Sports Med 2007;41(7):415–9.
29. Roseman T, Backenstrass M, Rosemann A, et al. Predictors of depression in a sample of 1021 primary care patients with osteoarthritis. Arthritis Rheum 2007;57(3):415–22.
30. Wise BL, Niu J, Zhang Y, et al. Psychological factors and their relation to osteoarthritis pain. Osteoarthritis Cartilage 2010;18:883–7.
31. Lin EH, Katon W, Von Korff M, et al. Effect of improving depression care on pain and functional outcomes among older adults with arthritis: a randomized controlled trial. JAMA 2003;290(18):2428–9.
32. Weigl M, Angst F, Aeschlimann A, et al. Predictors for response to rehabilitation in patients with hip or knee osteoarthritis: a comparison of logistic regression models with three different definitions of responder. Osteoarthritis Cartilage 2006;14(7):641–51.
33. Jones A, Doherty M. Intra-articular corticosteroids are effective in osteoarthritis but there are no clinical predictors of response. Ann Rheum Dis 1996;55:829–32.
34. Walker J, Sheather-Reid R, Carmody J, et al. Nonsteroidal antiinflammatory drugs in rheumatoid arthritis and osteoarthritis. Arthritis Rheum 1997;40(11):1944–54.
35. Palmieri-Smith RM, Thomas AC, Karvonen-Gutierrez C, et al. Isometric quadriceps strength in women with mild, moderate, and severe knee osteoarthritis. Am J Phys Med Rehabil 2010;89(7):541–8.

36. Segal NA, Torner JC, Felson D, et al. Effect of thigh strength on incident radiographic and symptomatic knee osteoarthritis in a longitudinal cohort. Arthritis Rheum 2009;61(9):1210–7.
37. Amin S, Baker K, Niu J, et al. Quadriceps strength and the risk of cartilage loss and symptom progression in knee osteoarthritis. Arthritis Rheum 2009;60(1): 189–98.
38. Roos EM, Herzog W, Block JA, et al. Muscle weakness, afferent sensory dysfunction and exercise in knee osteoarthritis. Nat Rev Rheumatol 2012;7(1):57–63.
39. Suetta C, Aagaard P, Magnusson S, et al. Muscle size, neuromuscular activation, and rapid force characteristics in elderly men and women: effects of unilateral long-term disuse due to hip osteoarthritis. J Appl Physiol 2007;102:942–8.
40. Fransen M, McConnell S. Land-based exercise for osteoarthritis of the knee: a metaanalysis of randomized controlled trials. J Rheumatol 2009;36(6):1109–17.
41. Fransen M, McConnell S, Hernandez-Molina G, et al. Does land-based exercise reduce pain and disability associated with hip osteoarthritis? A meta-analysis of randomized controlled trials. Osteoarthritis Cartilage 2010;18(5):613–20.
42. Wright A, Cook C, Flynn T, et al. Predictors of response to physical therapy intervention in patients with primary hip osteoarthritis. Phys Ther 2011;91(4):510–24.
43. Sattler M, Dannhauer T, Hudelmaier M, et al. Side differences of thigh muscle cross-sectional areas and maximal isometric muscle force in bilateral knees with the same radiographic disease stage, but unilateral frequent pain—data from the osteoarthritis initiative. Osteoarthritis Cartilage 2012;20(6):532–40.
44. Pietrosimone BG, Hertel J, Ingersoll CD, et al. Voluntary quadriceps activation deficits in patients with tibiofemoral osteoarthritis: a meta-analysis. PM R 2011; 3(2):153–62 [quiz: 62].
45. Scopaz KA, Piva SR, Gil AB, et al. Effect of baseline quadriceps activation on changes in quadriceps strength after exercise therapy in subjects with knee osteoarthritis. Arthritis Rheum 2009;61(7):951–7.
46. Gudbergsen H, Boesen M, Lohmander LS, et al. Weight loss is effective for symptomatic relief in obese subjects with knee osteoarthritis independently of joint damage severity assessed by high-field MRI and radiography. Osteoarthritis Cartilage 2012;20(6):495–502.
47. Hunter DJ, Niu J, Felson DT, et al. Knee alignment does not predict incident osteoarthritis: the Framingham osteoarthritis study. Arthritis Rheum 2007;56(4): 1212–8.
48. Brouwer G, vanTol A, Bergink A, et al. Association between valgus and varus alignment and the development and progression of radiographic osteoarthritis of the knee. Arthritis Rheum 2007;56(4):1204–11.
49. Hinman RS, Payne C, Metcalf BR, et al. Lateral wedges in knee osteoarthritis: what are their immediate clinical and biomechanical effects and can these predict a three-month clinical outcome? Arthritis Rheum 2008;59(3):408–15.
50. Cooke T, Sled E, Scudamore R. Frontal plane knee alignment: a call for standardised measurement. J Rheumatol 2007;34:1796–801.
51. Andriacchi T. Dynamics of knee malalignment. Orthop Clin North Am 1994;25: 395–403.
52. Bennell KL, Hunt MA, Wrigley TV, et al. Hip strengthening reduces symptoms but not knee load in people with medial knee osteoarthritis and varus malalignment: a randomised controlled trial. Osteoarthritis Cartilage 2010;18(5):621–8.
53. Foroughi N, Smith RM, Lange AK, et al. Progressive resistance training and dynamic alignment in osteoarthritis: a single-blind randomised controlled trial. Clin Biomech (Bristol, Avon) 2011;26(1):71–7.

54. Sharma L, Dunlop D, Cahue S, et al. Quadriceps strength and osteoarthritis progression in malaligned and lax knees. Ann Intern Med 2003;138:613–9.
55. Sharma L, Song J, Felson D, et al. The role of knee alignment in disease progression and functional decline in knee osteoarthritis. JAMA 2001;286:792.
56. Lim BW, Hinman RS, Wrigley TV, et al. Varus malalignment and its association with impairments and functional limitations in medial knee osteoarthritis. Arthritis Rheum 2008;59(7):935–42.
57. Lim BW, Hinman RS, Wrigley TV, et al. Does knee malalignment mediate the effects of quadriceps strengthening on knee adduction moment, pain and function in medial knee osteoarthritis? A randomized controlled trial. Arthritis Rheum 2008;59(7):943–51.
58. Guermazi A, Burstein D, Conaghan P, et al. Imaging in osteoarthritis. Rheum Dis Clin North Am 2008;34(3):645–87.
59. Conaghan PG, Hunter DJ, Maillefert JF, et al. Summary and recommendations of the OARSI FDA osteoarthritis Assessment of Structural Change Working Group. Osteoarthritis Cartilage 2011;19(5):606–10.
60. Gudbergsen H, Boesen M, Christensen R, et al. Radiographs and low field MRI (0.2T) as predictors of efficacy in a weight loss trial in obese women with knee osteoarthritis. BMC Musculoskelet Disord 2011;12:56.
61. Case JP, Baliunas AJ, Block JA. Lack of efficacy of acetaminophen in treating symptomatic knee osteoarthritis: a randomized, double-blind, placebo-controlled comparison trial with diclofenac sodium. Arch Intern Med 2003;163(2):169–78.
62. Deshmukh AJ, Panagopoulos G, Alizadeh A, et al. Intra-articular hip injection: does pain relief correlate with radiographic severity of osteoarthritis? Skeletal Radiol 2011;40(11):1449–54.
63. Plant M, Borg A, Dziedzic K, et al. Radiographic patterns and response to corticosteroid hip injection. Ann Rheum Dis 1997;56:476–80.
64. Atchia I, Kane D, Reed M, et al. Efficacy of a single ultrasound-guided injection for the treatment of hip osteoarthritis. Ann Rheum Dis 2011;70:110–6.
65. Loeser RF, Goldring SR, Scanzello CR, et al. Osteoarthritis: a disease of the joint as an organ. Arthritis Rheum 2012;64(6):1697–707.
66. Chao J, Wu C, Sun B, et al. Inflammatory characteristics on ultrasound predict poorer longterm response to intraarticular corticosteroid injections in knee osteoarthritis. J Rheumatol 2010;37(3):650–5.
67. Pendleton A, Millar A, O'Kane D, et al. Can sonography be used to predict the response to intra-articular corticosteroid injection in primary osteoarthritis of the knee? Scand J Rheumatol 2008;37(5):395–7.
68. Babyak M. What you see may not be what you get: a brief, non-technical introduction to overfitting in regression-type models. Psychosom Med 2004;66: 411–21.

Appendix 4: Does clinical presentation predict response to a non-surgical chronic disease management program for participants with end-stage hip and knee OA?

Does Clinical Presentation Predict Response to a Nonsurgical Chronic Disease Management Program for Endstage Hip and Knee Osteoarthritis?

Jillian P. Eyles, Barbara R. Lucas, Jillian A. Patterson, Matthew J. Williams, Kate Weeks, Marlene Fransen, and David J. Hunter

ABSTRACT. ***Objective.*** To identify baseline characteristics of participants who will respond favorably following 6 months of participation in a chronic disease management program for hip and knee osteoarthritis (OA).

Methods. This prospective cohort study assessed 559 participants at baseline and following 6 months of participation in the Osteoarthritis Chronic Care Program. Response was defined as the minimal clinically important difference of an 18% and 9-point absolute improvement in the Western Ontario and McMaster Universities Arthritis Index global score. Multivariate logistic regression modeling was used to identify predictors of response.

Results. Complete data were available for 308 participants. Those who withdrew within the study period were imputed as nonresponders. Three variables were independently associated with response: signal joint (knee vs hip), sex, and high level of comorbidity. Index joint and sex were significant in the multivariate model, but the model was not a sensitive predictor of response.

Conclusion. Strong predictors of response to a chronic disease management program for hip and knee OA were not identified. The significant predictors that were found should be considered in future studies. (First Release Sept 15 2014; J Rheumatol 2014;41:2223–31; doi:10.3899/jrheum.131475)

Key Indexing Terms:

OSTEOARTHRITIS	CHRONIC DISEASE	REHABILITATION
MULTIDISCIPLINARY	PREDICTOR	REGRESSION ANALYSIS

Osteoarthritis (OA) is one of the world's top 10 most disabling conditions[1]. According to global burden of disease estimates, musculoskeletal (MSK) disorders rank second only to mental and behavioral disorders in overall contribution to years lived with a disability (YLD)[2]. A large proportion of YLD attributed to MSK disorders results from hip and knee OA, estimated at over 17 million YLD worldwide[2].

From the Physiotherapy Department and Department of Rheumatology, Royal North Shore Hospital; Northern Clinical School; Kolling Institute of Medical Research; Institute of Bone and Joint Research; University of Sydney, Sydney; Physiotherapy Department, Wollongong Hospital, Wollongong, New South Wales, Australia.

J.P. Eyles, BAppSc(Phty); B.R. Lucas, MPH, FACP, BAppSc(Phty), Physiotherapy Department, Royal North Shore Hospital; J.A. Patterson, BScAdv(Hons), MBiostat, Kolling Institute of Medical Research; M.J. Williams, BAppSc(Phty), Physiotherapy Department, Royal North Shore Hospital; K. Weeks, BAppSc(Phty), Physiotherapy Department, Wollongong Hospital; M. Fransen, PhD, MPH, University of Sydney; D.J. Hunter, MBBS, PhD, FRACP, Department of Rheumatology, Royal North Shore Hospital; Northern Clinical School, Kolling Institute of Medical Research, Institute of Bone and Joint Research, University of Sydney.

Address correspondence to Prof. D.J. Hunter, Department of Rheumatology, Royal North Shore Hospital, Reserve Road, St. Leonards, Sydney, New South Wales, Australia. E-mail: David.Hunter@sydney.edu.au

Accepted for publication July 15, 2014.

Treatments for this disabling, prevalent, and incurable disease focus on symptomatic relief. Numerous international evidence-based guidelines for management of hip and knee OA have become available[3,4,5,6,7,8,9]. There is consistency in most of the recommendations made by the guidelines[10] and agreement that nonsurgical management of hip and knee OA should combine both nonpharmacological and pharmacological treatment modalities[3,4,5,6,7,8,9]. However, the recommendations are numerous and are not arranged systematically to indicate the order of priority in which treatments should be undertaken or which combinations of modalities should be used. Faced with a plethora of choices, it would be helpful for clinicians to be able to base treatment decisions on the identification of specific clinical presentations that foretell greater likelihood of success following implementation of an individual or combination of treatments. In an era when the delivery of quality care is being promoted coupled with finite resources, the ability to predict outcome/s to intervention would allow clinicians to prioritize those who will get the greatest benefit.

There is a growing body of evidence for clinical characteristics that predict response to nonsurgical interventions for participants with hip and knee OA[11]. Four previous studies attempted to identify predictors of response to

programs involving combinations of nonsurgical interventions[12,13,14,15]; however, consistent predictors of response were not found. All 4 treatment protocols involved strategies for self-management of OA including dietary advice; 2 studies provided weight loss advice if indicated[13,14], and all were of relatively short duration, ranging from 3 to 12 weeks[12,13,14,15]. To our knowledge, studies reporting outcomes or predictors of response to longer duration self-management programs do not exist. The aim of our research was to determine participant characteristics predictive of favorable outcomes following participation in a longer-term nonsurgical chronic disease management program for hip and knee OA. We hypothesized that it would be possible to predict participants likely to respond to the program using baseline demographic, psychological, disease-related, and functional performance variables.

MATERIALS AND METHODS

Study design. This observational clinical cohort study followed consecutive participants of the Osteoarthritis Chronic Care Program (OACCP) from 2 teaching hospitals in New South Wales (NSW), Australia, over a period of 6 months. The OACCP was developed by the Agency for Clinical Innovation MSK Network in response to the growing recognition of the need for a nonsurgical care program for people awaiting elective hip or knee joint replacement surgery (JRS). Participants with symptomatic and radiographic hip and knee OA were recruited for the OACCP at Royal North Shore/Ryde and Wollongong Hospitals from JRS waiting lists or referral by rheumatologists, orthopedic surgeons, and general practitioners. This equates to a doctor diagnosis of OA, which provides good face validity[16]. People with a diagnosis of knee or hip OA were eligible for the OACCP at initial assessment if they had pain in the affected knee/hip on most days of the past month[17]. Participants who had completed a reassessment at 26 weeks (within 140–225 days following initial assessment) were included in the analysis (Figure 1). There were no exclusion criteria for the OACCP, but participants who did not return for their 26-week assessment, or who were reassessed outside 140–225 days following initial assessment, were considered for imputation as nonresponders. Participants imputed as nonresponders included those who underwent JRS more than 90 days (and less than 225 days) following initial assessment, those discharged on medical advice, or participants who cited dissatisfaction with the program as their reason for discharge. Those receiving JRS within 90 days of initial assessment were excluded from analysis on the basis that there was insufficient time to determine whether they responded to the OACCP. The remaining participants without a complete 26-week assessment within 140-225 days were excluded from the regression analysis.

Intervention. The OACCP aimed to reduce pain, increase function, and improve the quality of life of participants with hip and knee OA through provision of access to relevant health professionals to support self-management and longterm behavior change. At initial assessment, the MSK Coordinator engaged participants to set goals around the management of their OA and comorbidities[18]. The MSK Coordinator was a specialized MSK physical therapist; all participants were prescribed an individualized exercise program that focused on strengthening muscles around affected joints, increasing physical activity levels, and other exercises depending on clinical presentation. These programs were reviewed at 12 and 26 weeks. All participants were provided with education about their OA and any identified comorbidities.

Following initial assessment, participants were referred to members of the multidisciplinary team (MDT) according to clinical need. If participants required medication review they were referred to a rheumatologist or pain clinical nurse consultant. Intraarticular injections were not part of the treatment provided. A dietitian provided interventions when indicated to assist participants with weight loss and/or comorbidity management. Participants requiring assessment of efficiency and safety of functional tasks were referred to an occupational therapist. Psychosocial interventions and linkage with community support services were provided by a social worker as required. Some participants with tibiofemoral or patellofemoral joint malalignment were referred to an orthotist for application of knee bracing or foot orthoses. Participants were also referred to healthcare providers outside the MDT for other interventions (e.g., hydrotherapy, diabetes education, psychology, etc.) as required.

Outcome measures. During a structured interview at initial assessment, the MSK coordinator recorded demographic and comorbidity data. Demographic data included sex, date of birth, referral source, residential status, language spoken at home, employment, and level of education. Signal joint, the predominant site of OA, was determined by clinical examination, patients' symptoms, and radiographic evidence of disease. All physical measures performed at initial and 26-week assessments were performed using a standardized protocol[17], including height, weight, waist and hip circumferences, and body mass index (BMI). Disease-specific self-report measures administered at 0 and 26 weeks included the Hip Dysfunction and Osteoarthritis Score (HOOS) or Knee Injury and Osteoarthritis Score (KOOS), according to the signal joint. The Depression Anxiety Stress 21 Scale (DASS 21) was used to measure these 3 negative emotional states at initial and 26-week assessments. The Six-minute Walk Test (6MWT) was completed at baseline and 26 weeks.

The validated, disease-specific HOOS[19] and KOOS[20] require participants to use 5-point Likert scales to rate their symptoms, stiffness, pain, physical function, recreational activities, and quality of life. The HOOS and KOOS subsume all of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) questions, enabling conversion into WOMAC scores[21,22]. The WOMAC subscales for pain, stiffness, and function were calculated by summation of the numerical responses provided by the WOMAC questions within the HOOS and KOOS. The WOMAC subscores were combined to calculate a WOMAC global score = 100– (sum of pain + stiffness + function items) × 100/96. Normalized WOMAC global scores were used, reflecting the convention that 100 indicated no problems and 0 indicated severe problems[21,22].

Using a 4-point Likert scale, the DASS 21 asks participants to rate how much 21 separate statements applied to them over the past week. The DASS 21 provides subscores to indicate the presence or absence of symptoms of depression, anxiety, and stress and has previously been shown to predict the diagnostic presence of depression and anxiety in older adults[23]. We were concerned primarily with the depression subscores; with 0–9 indicating no depressive symptoms, 10–13 mild, 14–20 moderate, 21–27 severe, and greater than 28, extremely severe symptoms. The DASS depression subscores were categorized into low depressive (0–13) versus high depressive (≥ 14) groups for the regression analyses.

The Modified Self-Administered Comorbidity Questionnaire asks participants, "Has your doctor told you that you have any of the following problems?" and then lists 21 commonly reported conditions plus an "other" category to indicate comorbidities not included on the list. Response is "yes" or "no". This questionnaire was adapted from The Self-Administered Comorbidity Questionnaire[24] and is scored by counting "yes" responses to indicate the number of comorbidities experienced by the participant. The number of comorbidities variable was categorized into low (0–2), high (3–5), and very high (≥ 6) groups for the analyses.

The 6MWT is recommended by the Osteoarthritis Research Society International to assess long-distance walking and aerobic capacity for participants with hip and knee OA[25]. Participants were asked to walk as quickly as they could for 6 min on a flat 25-m track with no corners[26] and the distance walked was recorded in meters. Baseline measurement of oxygen saturation, heart rate, and perceived exertion (Borg Scale) were taken prior to and at test completion. Participants with respiratory or cardiac concerns had measures taken at 1-min intervals during the test,

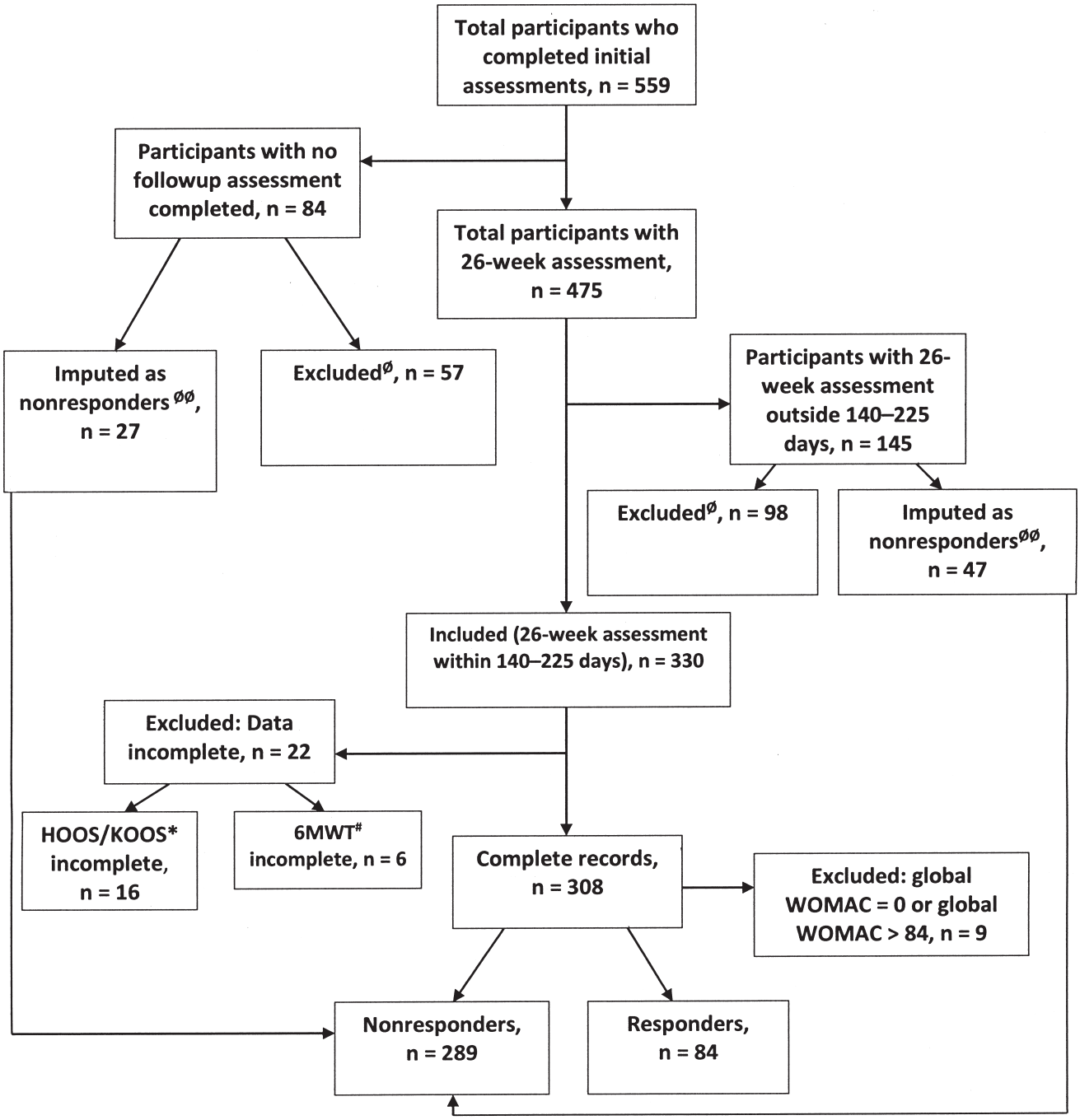


Figure 1. Study flowchart. Ø Participants with incomplete 26-week assessment or 26-week assessment outside 140–225 days or receiving joint replacement surgery (JRS) within 90 days of initial assessment. ØØ Participants who underwent JRS more than 90 days (and less than 225 days) following initial assessment, or were discharged on medical advice or who cited dissatisfaction with the program as the reason for their discharge. *HOOS or KOOS at either 0 or 26 weeks were incomplete so that WOMAC global scores could not be calculated. #6MWT results were unavailable because participants were unable to complete the test: 5 because of high blood pressure and 1 with back pain. HOOS: Hip Dysfunction and Osteoarthritis Score; KOOS: Knee Injury and Osteoarthritis Score; WOMAC: Western Ontario and McMaster Universities Arthritis Index; 6MWT: Six-minute Walk Test.

which was discontinued for the following: chest pain or discomfort, mental confusion, lack of coordination, dizziness, intolerable dyspnea, leg cramps, extreme muscle fatigue, persistent oxygen saturation < 85%, or other clinically warranted reasons.

Participants were asked to rate their average pain on the day of assessment using a visual analog scale (VAS; 0 indicated no pain and 10 the most pain imaginable). The pain VAS was categorized into low pain (VAS 0–5) and high pain (VAS 6–10) for the regression analyses.

Statistical analyses. Participants were dichotomized according to response or non-response at the 26-week assessment according to treatment based on

the notion of minimal clinically important difference (MCID), which can be defined as the smallest difference in scores of the variable concerned that is considered beneficial by participants of the intervention[27]. The MCID used was first developed by Angst, *et al*[28] to reflect the treatment effect considered to be clinically relevant to a comprehensive rehabilitation intervention for participants with OA of the lower extremities. This MCID required a relative change greater than or equal to 18% (100 × change of score/baseline score) and an absolute change of 9 points improvement of WOMAC global scores at the 26-week assessment compared to baseline. Using an MCID comprising both relative and absolute change standardized the amount of improvement required to achieve response across the spectrum of disease severity. Hence participants with very low global WOMAC scores were not classified as responders for small absolute changes in score compared with those whose baseline scores were higher. Participants who demonstrated improvements in WOMAC global scores at 26 weeks of ≥ 18% with an absolute change in score ≥ 9 were categorized as responders[28]; those who did not were nonresponders.

Participants censored at their 26-week followup because of JRS performed at least 90 days after their initial assessment and within the 26-week assessment window (≤ 225 days) were imputed into the analysis as nonresponders. Participants who withdrew from the OACCP owing to dissatisfaction with the program or following medical advice were also imputed as nonresponders.

The potential predictor variables were chosen following literature review[11] and discussion among this study's authors. The MSK coordinators collecting the data at both study sites were blinded to which variables were to be analyzed as predictors of response. Eight baseline predictor variables were identified *a priori* for consideration in the model: BMI, pain VAS, DASS subscore, signal joint, 6MWT, age, sex, and number of comorbidities. The power calculation was set to include at least 10 "responders" per predictor variable[29,30]. Previous studies have reported 34%–47% of participants with hip or knee OA may be expected to satisfy responder criteria following nonsurgical multimodal interventions[14,15]. A sample of 267 was considered sufficient to accommodate 8 predictor variables.

Univariate logistic regression analyses examined the association between each of the predictor variables and response, and continuous variables were categorized when necessary to meet linearity requirements. All variables were entered into a multivariate binary logistic regression model; the least significant predictor was removed at each step of the modeling until only significant variables remained. To control for confounding, when any variables associated with response in the univariate analyses were removed from the model, the regression coefficients of the remaining variables were checked for a change in 10% or more and if so were retained. Testing for interactions was performed by combining variables of interest. SPSS version 21 was used for all statistical analyses.

Ethics approval was granted by the NSW Population and Health Services Research Ethics Committee (AUREI Reference HREC/12/CIPHS/63); Cancer Institute NSW Reference Number 2012/08/413.

RESULTS

Of 559 patients consecutively referred to the Wollongong and Royal North Shore/Ryde Hospitals OACCP from July 2011 to December 2013, 475 participants had completed their 26-week assessment as shown in Figure 1. There were 145 participants who were excluded because their 26-week assessment occurred outside the assessment range. A further 16 participants were excluded with incomplete HOOS or KOOS, 6 were unable to complete the 6MWT because of high blood pressure or back pain, and 84 did not return for followup assessment. That left 308 participants with complete datasets remaining for the analysis. A further 74 were imputed as nonresponders: 55 discharged from the OACCP after JRS 90–225 days following initial assessment, 16 withdrew because of dissatisfaction with the program, and 3 stopped as a result of medical advice.

The baseline demographics of included participants, those excluded because of missing assessments or assessments outside the 26-week range (n = 167), and those who did not return for followup assessment (n = 84) are summarized in Table 1. The included and excluded groups were homogeneous in most respects. About 90% were referred from elective JRS waiting lists; the wait time for JRS in NSW Hospitals is around 12 months. The majority of participants were of similar age, lived at home with an able person, spoke English, were retired, and overweight. Participants reported similar baseline pain. The majority had 0–5 comorbidities and did not finish high school.

There were proportionally more males in the excluded group with no followup assessment ($p = 0.07$) and the included group reported a higher proportion of OA knees to hips than did the excluded groups ($p = 0.02$). The mean baseline WOMAC global scores were significantly different ($p = 0.03$); however, the greatest difference in mean scores was 5.2 points, which is not very clinically important.

The referrals to healthcare providers recorded for included and excluded participants are summarized in Table 2. All participants were assessed by a physical therapist and provided with a graded exercise program; around half were referred to a dietitian; 30-40% to a rheumatologist; and 20–30% to an occupational therapist or a social worker. About 20% of participants were referred to providers within and 40% outside the local health district.

Of 308 included participants with complete datasets, 9 were omitted from analysis because their baseline WOMAC was too high (> 84) or 0, and so were unable to achieve response. Of the 299 participants with complete datasets, 84 (28%) were responders according to the MCID. Results of the univariate regression analyses are shown in Table 3. Compared to females, males were less likely to be responders (OR 0.5, 95% CI 0.31, 0.88). There was strong evidence that participants with knee OA were more likely to be responders than those with hip OA (OR 2.1, 95% CI 1.10, 3.88). Compared to those with a low number (≤ 2) there was evidence that participants with a very high number of comorbidities (≥ 6) were more likely to be responders (OR 2.2, 95% CI 0.99, 4.95). The other baseline variables were not independently associated with response.

All potential predictor variables were entered into the base multivariate model. No significant interactions between the variables were found. Following elimination of nonsignificant variables, the final model (Table 4) contained both signal joint (chi $\text{square}_{LR} = 4.49$, $p < 0.05$) and sex (chi $\text{square}_{LR} = 4.95$, $p < 0.05$). Participants with the knee as the signal joint were more likely to be responders compared with those with hip as the signal joint (adjusted OR 1.92, 95% CI 1.02, 3.62). Compared to women, men were less

Table 1. Demographics of included and excluded participants at baseline.

Baseline Characteristics	Included, n = 313	Excluded: 26 Weeks Not Within 140–225 Days, or Missing Data, n = 162	Excluded: No Followup Assessment, n = 84	p [v]
Female (%)	62	59	48	0.067
Age, yrs, mean (SD)	68.5 (9.25)	69.0 (9.92)	68.0 (10.85)	0.76
Signal joint knee (%)	77	65	68	**0.022**
Signal joint knee (%) responders	83			
Signal joint knee (%) nonresponders	75			
On elective joint replacement list (%)	88	90	86	0.68
Residence				
At home with able person (%)	64	68	68	0.46
Home alone (%)	28	22	21	
Other[‡] (%)	8	10	11	
Speaks English[*] (%)	90	92	88	0.61
Employment (%)				0.60
Not currently employed[†]	86	82	84	
Currently employed[^]	14	18	16	
Education (%)				0.94
Finished secondary school or higher[¤]	30	29	32	
Did not finish secondary school[°]	60	71	68	
BMI, mean (SD)	31.9 (6.88)	32.0 (6.57)	31.7 (6.36)	0.94
BMI knees, mean (SD)	32.52 (7.12)			
BMI hips, mean (SD)	30.03 (5.84)			
Pain VAS, mean (SD)	5.5 (1.84)	5.7 (1.74)	5.7 (2.20)	0.65
No. comorbidities (%)				
Low (0–2)	54	44	43	
High (3–5)	39	51	42	
Very high (≥ 6)	8	5	10	
Missing (no.)			5	
WOMAC global score[#], mean (SD), range	43.4 (19.39), 0–100	38.4 (17.17), 0–90	41.3 (21.72), 3–98	**0.027**
WOMAC global score, knees, mean (SD)	44.2 (19.66)	40.8 (18.59)	41.5 (21.23)	
WOMAC global score, hips, mean (SD)	40.7 (18.34)	33.6 (12.96)	40.8 (23.23)	
WOMAC global score for responders, mean (SD), range	33.8 (18.06), 1–79			
WOMAC global score for nonresponders, mean (SD), range	47.4 (18.51), 4–100			
6 Minute Walk Test, m, mean (SD)	337.4 (118.52)	324.3 (120.51)	323.5 (114.51)	0.44

Data in bold face are statistically significant. [‡] Other includes residence at hostel or residence with non-able person. * Participants who did not speak English (about 10%) required the use of an interpreter. [†] Not currently employed includes participants who reported they were retired, performed home duties, and other. [^] Currently employed includes participants who reported engaging in full/part time/volunteer work. [¤] Includes participants who reported finishing secondary school, tertiary certificate, or university graduate. [°] Includes participants who did not finish secondary school, and those who reported no formal schooling. [#] The WOMAC global scores are a transformed score calculated from the HOOS and KOOS: 100 indicates no problems and 0 indicates extreme problems. [v] Independent ANOVA or chi-squared statistic comparing included participants with the 2 other groups. VAS: visual analog scale; WOMAC: Western Ontario and McMaster Universities Arthritis Index; BMI: body mass index.

likely to be responders (adjusted OR 0.55, 95% CI 0.32, 0.94). The group with a very high number of comorbidities was not significantly associated with response in the multivariate model (p = 0.07) and removal did not have a confounding effect on the remaining variables. The model fit the data well using the Hosmer-Lemeshow goodness-of-fit test (chi square = 3.03, 3 DF, p = 0.21); however, the model was unable to predict any participants as responders (sensitivity 0%, specificity 100%).

DISCUSSION

To our knowledge, this was the first study attempting to identify predictors of response following longer-term (6 mos) participation in a chronic disease management

Table 2. Percentage of OACCP participants referred to health providers within and outside the OACCP.

Healthcare Provider Type	Included Participants, n = 300	Participants Without Followup, Imputed as Nonresponders, n = 74	Excluded Participants, n = 185
OACCP multidisciplinary team			
OACCP physical therapist* (%)	100	100	100
OACCP dietitian* (%)	53.7	55.7	41.3
OACCP rheumatologist# (%)	40.4	46.3	31.1
OACCP occupational therapist* (%)	28.4	36.6	30.5
OACCP social worker# (%)	19.4	28.5	13.8
OACCP orthotist# (%)	23.7	17.8	13.5
Other** (%)	16	12.9	10.7
Other health providers within the local health district± (e.g., hydrotherapy, exercise groups; %)	21	20.3	19.4
Other health providers outside the local health district‡ (e.g., GP, hydrotherapy, diabetes educator, exercise groups; %)	42	39.2	39.2

* Available at both OACCP sites. # Available at Royal North Shore Hospital OACCP (only this rheumatologist saw patients in the OACCP clinic, they did not refer participants to the OACCP). **Other may include pain CNC at Wollongong Hospital and education sessions at both sites. ± Other healthcare providers within the local health district may include hydrotherapy, exercise groups, falls clinic, physiotherapist, pulmonary rehabilitation, smoking cessation, or geriatrician. ‡Other healthcare providers outside the local health district may include general practitioner (GP), hydrotherapy, exercise groups, diabetes clinic, orthopedic surgeon, psychologist, geriatrician, physiotherapist, dietitian, falls clinic, pain clinic, social worker, orthotist, smoking cessation, pulmonary and cardiac rehabilitation. OACCP: Osteoarthritis Chronic Care Program; CNC: clinical nurse consultant.

program for hip and knee OA. The relatively low response rate (28%) was not surprising considering the severity of disease in this sample indicated by the large proportion of participants on JRS waiting lists (around 90%). Assuming that participants on JRS waiting lists would have clinically and radiographically significant disease, it may be expected that given the natural history of the disease, without intervention the majority of participants would stay the same or worsen over a period of 6 months. A similar response rate was reported by Weigl, *et al* using a less stringent definition of response (≥ 18% improvement in global WOMAC score) 6 months following a 3–4 week rehabilitation program for participants with hip and knee OA[15].

The univariate analysis and the multivariate model adjusting for sex found participants with the knee as signal joint were almost twice as likely to be responders compared to those referred with hip OA (OR 1.92, 95% CI 1.02, 3.62). Although signal joint is not a significant predictor of response in the literature[13,14,15], this finding makes sense in the clinic. A central aim of the OACCP was to increase physical activity. There is evidence that participants with knee OA experience reduced pain and improvement in physical function following land-based therapeutic exercise[31]; however, the evidence for such benefits is weaker in those with hip OA[32]. Perhaps the participants with knee OA derived higher levels of therapeutic benefit from the exercise prescribed by the physical therapist of the OACCP and so were more likely to respond than were those with hip OA. Included participants with knee OA had a higher mean BMI (32.52 kg/m^2) than those with hip OA (30.03 kg/m^2; Table 1). Given that a common goal for OACCP participants was to lose weight, and that participants with knee OA were more overweight, it was hypothesized that knees would be more likely to respond to interventions that involved weight loss. Interestingly, BMI was not an independent predictor of response, and it was not significant in the multivariate model when adjusted for signal joint. This confirms previous findings that BMI was not predictive of responsiveness to weight loss or multimodal nonpharmacological and pharmacological interventions for participants with hip and knee OA[13,14,33].

Sex was a univariate predictor of response that remained significant in the multivariate model adjusting for signal joint (OR 0.54, 95% CI 0.31, 0.95). Men were half as likely to be responders as women, a result that is difficult to explain. The literature yields conflicting results: being female was predictive of response to a rehabilitation program for hip and knee OA[15], but sex was not significantly associated with response in other previous predictor studies[13,14].

Compared to participants with a low number of comorbidities (0–2), participants with a very high number of comorbidities (> 6) were independently associated with response (OR 2.2, 95% CI 0.99, 4.95). A very high number of comorbidities was not significantly associated with response when adjusting for sex and signal joint, so number of comorbidities was removed from the model.

The absence of depression has been identified previously as a predictor of response to a 3–4 week inpatient multi-

Table 3. Univariate analyses of potential predictors of response to the OACCP.

Variable		Unadjusted OR (95% CI)	p
Age		0.9 (0.071, 1.20)	0.539
Sex	Female	Reference	
	Male	0.5 (0.31, 0.88)	0.015
Signal joint	Knee	2.1 (1.10, 3.88)	0.023
	Hip	Reference	
Comorbidity	Low (0–2)	Reference	
	High (3–5)	0.8 (0.47, 1.37)	0.414
	Very high (≥ 6)	2.2 (0.99, 4.95)	0.053
Depression*	≤ 13	Reference	
	≥ 14	1.2 (0.68, 1.98)	0.592
Pain †	0–5	Reference	
	6–10	1.2 (0.72, 1.92)	0.526
BMI		1.0 (0.98, 1.05)	0.329
6MWT**		1.0 (1.0, 1.0)	0.755

*Depression measured using the Depression component of the Depression Anxiety Stress Scales. †Pain measured using visual analog scale (self-rated; 0 no pain, 10 worst pain). **Distance participants are able to walk on flat ground during Six-minute Walk Test. OACCP: Osteoarthritis Chronic Care Program; BMI: body mass index.

Table 4. Final multivariate[#] prediction model for response to the Osteoarthritis Chronic Care Program.

Variable	β Coefficient	p	Adjusted OR	95% CI
Constant	–1.496			
Sex	–0.594	0.029	0.55	0.32, 0.94
Signal joint knee	0.651	0.045	1.92	1.02, 3.62

[#] The base adjusted or multivariate model included age, sex, index joint, comorbidity, depression, pain, body mass index, and Six-minute Walk Test.

modal rehabilitation intervention[15] and positive outcomes from a weight loss program in overweight veterans with knee OA[33]. The absence of depression was not a significant predictor of response in the present study. Participants reporting depressive symptoms on the DASS depression subscale were referred for treatment as required. The treatment of depression in people with arthritis has been shown to reduce pain and depressive symptoms, and improve function and quality of life[34]. The treatment of depression as an adjunct to the other multidisciplinary interventions in our study may have diminished the negative effect depressive symptoms had on response to treatment.

Age was not a predictor of response to the OACCP. Most studies include age in their list of potential predictor variables to control for the effects of confounding. Previous evidence for age as a predictor of response is conflicting. Higher age was a predictor of response to a multimodal stepped-care model for participants with hip and knee OA[13] and a physical therapy intervention for patients with hip OA[35] but was insignificant in other predictor studies[14,15]. The 6MWT was not predictive of response and while functional performance measures have not been widely used in previous prediction studies, 1 study found the self-paced 40-m walk test predictive of response to physical therapy interventions for patients with hip OA[35]. A recent systematic review rated the 40-m walk test as the best walk test based on the limited evidence available[36] and perhaps it would have been a more useful predictor of response for our study. This is an interesting area for future research.

Notable strengths of this study design included the large sample size, the clinically meaningful followup period, and that the potential predictor variables were identified *a priori* through literature and peer review, with due consideration to not overfitting the model with excessive degrees of freedom. The potential predictors included a broad mix of disease, psychological, physical, and demographic variables. To minimize bias, the data were collected prospectively by the MSK coordinators, who were blinded to which variables were to be analyzed as predictors.

This clinical cohort study used data from a real-life clinic. The participants required doctor diagnosis of OA, which provides good face validity but may present potential limitations because different symptom labels for OA may exist between independent medical practitioners[16]. Recruited largely from JRS waiting lists, many participants of the OACCP were censored when their date for JRS came up. Excluded participants reported worse global WOMAC scores at baseline compared to included participants. To control for selection bias, participants who had experienced at least 90 days in the OACCP and had surgery within the 26-week window (≤ 225 days) were imputed as non-responders, in addition to those who discontinued the OACCP citing dissatisfaction with the program or who withdrew under medical advice. The transformed baseline global WOMAC score (100 indicates no problems and 0 indicates extreme problems) was significantly lower in responders compared to nonresponders ($p < 0.05$), and

although there is marked overlap between groups, the mean difference of 10 points may suggest some regression toward the mean.

A control group was not used in this study, so it could be argued that it is impossible to distinguish between predictors of response to the chronic disease management program and natural progression of the disease. Previous studies concerned with progression of OA indicate a slow evolution and progression of the disease over time[37]. Given that the vast majority of patients were on the waiting list for JRS indicating endstage disease, it would be unlikely that the natural course of OA in these participants would allow improvement in symptoms sufficient to achieve the MCID over a period of 6 months. However, this does limit the generalizability of the results of our study to those with severe OA. A previous study reported that compared to participants not waiting for surgery, patients on the waitlist for knee JRS experienced smaller improvements that were not as lasting in response to participation in a chronic disease management program[12]. It would be interesting in future research to investigate a more heterogeneous sample of participants to enable analysis of referral for JRS as a potential predictor of response.

We can only assume that referral for JRS was a proxy measure of disease severity in this study. Future research should include a standardized measure of structural disease severity. Higher radiographic severity of knee and hip OA measured using the Kellgren-Lawrence (KL) Grading Scale was a predictor of response to acetaminophen as part of a Dutch multimodal stepped-care model[13]. Conversely, an earlier study investigating predictors of response to the same intervention found that KL grade was not associated with a more stringent definition of response[14]. It would be interesting to investigate whether radiographic severity is associated with response to the longer-term chronic disease management program. Another predictor variable in the literature associated with response was history of previous nonsurgical interventions. Two studies reported history of previous nonsurgical therapies as associated with good response to rehabilitation programs for participants with hip or knee OA[12,15]. This should be addressed in future studies concerned with prediction of response to chronic disease management programs for hip and knee OA.

Response to intervention could not be predicted using the variables studied in this sample following 6 months of participation in the OACCP. Although significant predictors of response were identified, the model was not sensitive. The significant predictors of our study should be considered for future research, and alternative variables for investigation have been highlighted. It is possible that an alternative battery of variables could be more useful for prediction of response to this intervention. If response can be predicted, it may enable clinicians to better tailor management of hip and knee OA according to clinical presentation.

ACKNOWLEDGMENT

Data for this analysis were obtained with permission of the Agency for Clinical Innovation (ACI). The authors thank the ACI MSK Network, especially Mary Fien, for assistance with accessing data. We are very grateful to Gary Rolls for his continued support, and Victoria Ireland for assistance with Royal North Shore Hospital data collection.

REFERENCES

1. World Health Organisation. Chronic diseases and health promotion: chronic rheumatic conditions. [Internet. Accessed Aug 7, 2014.] Available from: www.who.int/chp/topics/rheumatic/en/
2. Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012;380:2163-96.
3. Zhang W, Moskowitz RW, Nuki G, Abramson S, Altman RD, Arden N, et al. OARSI recommendations for the management of hip and knee osteoarthritis, Part II: OARSI evidence-based, expert consensus guidelines. Osteoarthritis Cartilage 2008;16:137-62.
4. Zhang W, Doherty M, Arden NK, Bannwarth B, Bijlsma J, Gunther KP, et al. EULAR evidence based recommendations for the management of hip osteoarthritis: report of a task force of the EULAR Standing Committee for International Clinical Studies Including Therapeutics (ESCISIT). Ann Rheum Dis 2005; 64:669-81.
5. Pendleton A, Arden N, Dougados M, Doherty M, Bannwarth B, Bijlsma JW, et al. EULAR recommendations for the management of knee osteoarthritis: report of a task force of the Standing Committee for International Clinical Studies Including Therapeutic Trials (ESCISIT). Ann Rheum Dis 2000;59:936-44.
6. Royal Australian College of General Practitioners. Guideline for the non-surgical management of hip and knee osteoarthritis. July 2009. [Internet. Accessed Aug 7, 2014.] Available from: www.racgp.org.au/your-practice/guidelines/musculoskeletal/hipandkneeosteoarthritis/
7. Zhang W, Nuki G, Moskowitz RW, Abramson S, Altman RD, Arden NK, et al. OARSI recommendations for the management of hip and knee osteoarthritis: part III: Changes in evidence following systematic cumulative update of research published through January 2009. Osteoarthritis Cartilage 2010;18:476-99.
8. National Institute for Health and Care Excellence (UK). Osteoarthritis: the care and management of osteoarthritis in adults. 2008. [Internet. Accessed Aug 7, 2014.] Available from: http://guidance.nice.org.uk/CG59.
9. Hochberg MC, Altman RD, April KT, Benkhalti M, Guyatt G, McGowan J, et al. American College of Rheumatology 2012 recommendations for the use of nonpharmacologic and pharmacologic therapies in osteoarthritis of the hand, hip, and knee. Arthritis Care Res 2012;64:465-74.
10. Misso M, Pitt V, Jones K, Barnes H, Piterman L, Green S. Quality and consistency of clinical practice guidelines for diagnosis and management of osteoarthritis of the hip and knee: a descriptive overview of published guidelines. Med J Aust 2008;189:394-9.
11. Eyles J, Lucas B, Hunter DJ. Targeting care tailoring nonsurgical management according to clinical presentation. Rheum Dis Clin North Am 2013;39:213-33.
12. Lamb SE, Toye F, Barker KL. Chronic disease management programme in people with severe knee osteoarthritis: efficacy and moderators of response. Clin Rehabil 2008;22:169-78.
13. Snijders GF. Treatment outcomes of a Numeric Rating Scale (NRS)-guided pharmacological pain management strategy in symptomatic knee and hip osteoarthritis in daily clinical practice. Clin Exp Rheumatol 2012;30:164-70.
14. Snijders GF, den Broeder AA, van Riel PL, Straten VH, de Man

FH, van den Hoogen FH, et al; NOAC Study Group. Evidence-based tailored conservative treatment of knee and hip osteoarthritis: between knowing and doing. Scand J Rheumatol 2011;40:225-31.
15. Weigl M, Angst F, Aeschlimann A, Lehmann S, Stucki G. Predictors for response to rehabilitation in patients with hip or knee osteoarthritis: a comparison of logistic regression models with three different definitions of responder. Osteoarthritis Cartilage 2006;14:641-51.
16. Jordan K, Clarke AM, Symmons DP, Fleming D, Porcheret M, Kadam UT, et al. Measuring disease prevalence: a comparison of musculoskeletal disease using four general practice consultation databases. Br J Gen Pract 2007;57:7-14.
17. New South Wales Agency for Clinical Innovation. Osteoarthritis Chronic Care Program Site manual. 2013. [Internet. Accessed Aug 7, 2014.] Available from: www.aci.health.nsw.gov.au/__data/assets/pdf_file/0007/190987/ACI-Musculoskeletal-Network-Osteoarthritis-Chronic-Care-Program-Site-Manual.pdf#zoom=100
18. New South Wales Agency for Clinical Innovation. Musculoskeletal Network — Osteoarthritis Chronic Care Program Model of Care. 2012. Available from: www.aci.health.nsw.gov.au/__data/assets/pdf_file/0020/165305/Osteoarthritis-Chronic-Care-Program-Mode-of-Care.pdf#zoom=100
19. Thorborg K, Roos EM, Bartels EM, Petersen J, Hölmich P. Validity, reliability and responsiveness of patient-reported outcome questionnaires when assessing hip and groin disability: a systematic review. Br J Sports Med 2010;44:1186-96.
20. Collins NJ, Misra D, Felson DT, Crossley KM, Roos EM. Measures of knee function: International Knee Documentation Committee (IKDC) Subjective Knee Evaluation Form, Knee Injury and Osteoarthritis Outcome Score (KOOS), Knee Injury and Osteoarthritis Outcome Score Physical Function Short Form (KOOS-PS), Knee Outcome Survey Activities of Daily Living Scale (KOS-ADL), Lysholm Knee Scoring Scale, Oxford Knee Score (OKS), Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), Activity Rating Scale (ARS), and Tegner Activity Score (TAS). Arthritis Care Res 2011;63 Suppl 11:S208-28.
21. KOOS User's Guide. 2003. [Internet. Accessed Aug 7, 2014.] Available from: www.koos.nu/KOOSGuide2003.pdf.
22. HOOS User's Guide. 2003. [Internet. Accessed Aug 7, 2014.] Available from: www.liv.se/Global/Jobb,%20utbildning%20och%20forskning/Forskning/FoU-enheter/Division%20HHR/Utv%C3%A4rderings instrument/HOOS%20User's%20Guide%202003.pdf
23. Gloster AT, Rhoades HM, Novy D, Klotsche J, Senior A, Kunik M, et al. Psychometric properties of the Depression Anxiety and Stress Scale-21 in older primary care patients. J Affect Disord 2008;110:248-59.
24. Sangha O, Stucki G, Liang MH, Fossel AH, Katz JN. The Self-Administered Comorbidity Questionnaire: a new method to assess comorbidity for clinical and health services research. Arthritis Rheum 2003;49:156-63.
25. Dobson F, Hinman RS, Roos EM, Abbott JH, Stratford P, Davis AM, et al. OARSI recommended performance-based tests to assess physical function in people diagnosed with hip or knee osteoarthritis. Osteoarthritis Cartilage 2013;21:1042-52.
26. Bennell K, Dobson F, Hinman R. Measures of physical performance assessments: Self-Paced Walk Test (SPWT), Stair Climb Test (SCT), Six-Minute Walk Test (6MWT), Chair Stand Test (CST), Timed Up & Go (TUG), Sock Test, Lift and Carry Test (LCT), and Car Task. Arthritis Care Res 2011;63 Suppl 11:S350-70.
27. Jaeschke R, Singer J, Guyatt G. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials 1989;10:407-15.
28. Angst F, Aeschlimann A, Beat M, Stucki G. Minimal clinically important rehabilitation effects in patients with osteoarthritis of the lower extremities. J Rheumatol 2002;29:131-8.
29. Peduzzi P, Concato J, Holford T, Feinstein A. The importance of events per independent variable in multivariable analysis, II: accuracy and precision of regression estimates. J Clin Epidemiol 1995;48:1503-10.
30. Peduzzi P, Concato J, Kemper E, Holford T, Feinstein A. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol 1996;49:1373-9.
31. Fransen M, McConnell S. Land-based exercise for osteoarthritis of the knee: a metaanalysis of randomized controlled trials. J Rheumatol 2009;36:1109-17.
32. Fransen M, McConnell S, Hernandez-Molina G, Reichenbach S. Does land-based exercise reduce pain and disability associated with hip osteoarthritis? A meta-analysis of randomized controlled trials. Osteoarthritis Cartilage 2010;18:613-20.
33. Wolf S, Foley S, Budiman-Mak E, Moritz T, O'Connell S, Jelinek C, et al. Predictors of weight loss in overweight veterans with knee osteoarthritis who participated in a clinical trial. J Rehabil Res Dev 2010;47:171.
34. Lin EH, Katon W, Von Korff M, Tang L, Williams JW Jr., Kronke K, et al. Effect of improving depression care on pain and functional outcomes among older adults with arthritis: a randomized controlled trial. JAMA 2003;290:2428-9.
35. Wright AA, Cook CE, Flynn TW, Baxter GD, Abbott JH. Predictors of response to physical therapy intervention in patients with primary hip osteoarthritis. Phys Ther 2011;91:510-24.
36. Dobson F, Hinman RS, Hall M, Terwee CB, Roos EM, Bennell KL. Measurement properties of performance-based measures to assess physical function in hip and knee osteoarthritis: a systematic review. Osteoarthritis Cartilage 2012;20:1548-62.
37. Felson DT, Zhang Y, Hannan MT, Naimark A, Weissman BN, Aliabadi P, et al. The incidence and natural history of knee osteoarthritis in the elderly. The Framingham Osteoarthritis Study. Arthritis Rheum 1995;38:1500-5.

Appendix 5: Can we predict those with OA who worsen following a chronic disease management program?

Arthritis Care & Research
Vol. 68, No. 9, September 2016, pp 1268–1277
DOI 10.1002/acr.22836

ORIGINAL ARTICLE

Can We Predict Those With Osteoarthritis Who Will Worsen Following a Chronic Disease Management Program?

JILLIAN P. EYLES,[1] KATHRYN MILLS,[2] BARBARA R. LUCAS,[3] MATTHEW J. WILLIAMS,[3] JOANNA MAKOVEY,[1] LAURENCE TEOH,[4] AND DAVID J. HUNTER[1]

Objective. **To identify predictors of worsening symptoms and overall health of the treated hip or knee joint following 26 weeks of a nonsurgical chronic disease management program for hip and knee osteoarthritis (OA) and to examine the consistency of these predictors across 3 definitions of worsening.**

Methods. **This prospective cohort study followed 539 participants of the program for 26 weeks. The 3 definitions of worsening included symptomatic worsening based on change in the Western Ontario and McMaster Universities Osteoarthritis Index Global score (WOMAC-G) measuring pain, stiffness, and function; a transition scale that asked about overall health of the treated hip or knee joint; and a composite outcome including both. Multivariate logistic regression models were constructed for the 3 definitions of worsening.**

Results. **Complete data were available for 386 participants: mean age was 66.3 years, 69% were female, 85% reported knee joint pain as primary symptom (signal joint), 46% were waitlisted for total joint arthroplasty (TJA). TJA waitlist status, signal joint, 6-Minute Walk Test (6MWT), depressive symptoms, pain, and age were independently associated with at least 1 definition of worsening. TJA waitlist status and 6MWT remained in the multivariate models for the transition and composite definitions of worsening.**

Conclusion. **Participants reporting worsening on the transition scale did not consistently meet the WOMAC-G definition of worsening symptoms. TJA waitlist status was predictive of the composite definition of worsening, a trend apparent for the transition definition. However, variables that predict worsening remain largely unknown. Further research is required to direct comprehensive and targeted management of patients with hip and knee OA.**

INTRODUCTION

Osteoarthritis (OA) is a well-known cause of significant disability (1,2). Current evidence promotes tailored treatments combining nonpharmacologic and pharmacologic nonsurgical modalities for symptomatic management of knee and hip OA (3–5). It would be naive to assume that everybody will benefit from a similar program of nonsurgical interventions, hence it is important to identify participants likely to report symptomatic worsening despite "usual" nonsurgical regimens so that alternative therapeutic options may be considered.

Longitudinal studies have examined predictors of hip and knee OA progression, i.e., deterioration in radiographic features, symptoms, or progression to total joint arthroplasty (TJA) (6–10). Attempts have been made to identify thresholds of pain, function, and structural severity to provide a surrogate measure of need for TJA for use in clinical trials (11,12). Elusive thus far, these thresholds could be useful to triage participants of nonsurgical self-management programs so that those who may potentially derive more benefit from TJA may be escalated to surgery. Extant thresholds indicative of symptomatic worsening following a rehabilitation program for hip and knee OA (13)

Ms. Eyles' work was supported by a Ramsay Health Care Allied Health Scholarship and a Royal North Shore Hospital Staff Specialist Award.

[1]Jillian P. Eyles, BAppSci(Physio), Joanna Makovey, PhD, David J. Hunter, MBBS, PhD, FRACP: Royal North Shore Hospital, and Kolling Institute of Medical Research, Institute of Bone and Joint Research, University of Sydney, Sydney, New South Wales, Australia; [2]Kathryn Mills, PhD: Macquarie University, Sydney, New South Wales, Australia; [3]Barbara R. Lucas, MPH, Matthew J. Williams, BAppSci(Physio): Royal North Shore Hospital, Kolling Institute of Medical Research, Institute of Bone and Joint Research, University of Sydney, Sydney, New South Wales, Australia; [4]Laurence Teoh, BHB, MBChB, FRACP: North Shore Hospital, Takapuna, Auckland, New Zealand.

Address correspondence to Jillian P. Eyles, BAppSci(Physio), Department of Rheumatology, Royal North Shore Hospital, Reserve Road, St. Leonards, 2065 NSW, Australia. E-mail: jillian.eyles@sydney.edu.au.

Submitted for publication June 16, 2015; accepted in revised form January 5, 2016.

Significance & Innovations

- Three different definitions of worsening were applied to a cohort participating in a chronic disease management program for hip and knee osteoarthritis.
- Participants who were "moderately" or "much" worse on the transition scale were not consistently worse, according to the minimal important difference using the Western Ontario and McMaster Universities Osteoarthritis Index global (WOMAC-G) scores.
- Predictors of worsening were similar between the transition and composite definitions, but not the WOMAC-G definition of worsening.
- Total joint arthroplasty waitlist status was significantly associated with the composite definition of worsening; however, variables that predict worsening remain largely unknown.

have not been applied extensively in the literature. Yet the ideal threshold for "worsening" may be useful by 1) interpreting clinical findings of individual patients and 2) deriving predictors of worsening in groups of nonsurgical program participants. Predictors may be used to triage referrals for interventions unlikely to confer benefit, placing unnecessary burden on the health care system and patients.

In the absence of a robust clinical definition of "worsening," we have compared 2 definitions existing in the literature. The first is based on a threshold of change in the Western Ontario and McMaster Universities Osteoarthritis Index global score (WOMAC-G) derived from participants of a rehabilitation intervention (13). The second is the response to a transition question measuring the overall amount and direction of change the individual has undergone regarding their joint following the intervention (14). This method has been suggested following recent questioning of the utility of pre- and post–patient-reported outcomes (PROs) to measure efficacy of self-management education programs, possibly due to a response shift prompted by change in the participant's perspective following engagement in such programs (14). A third definition of worsening was used in this study, i.e., a composite outcome including either or both WOMAC-G and transition definitions. The composite outcome was chosen to reflect the dual importance of the 2 clinical outcomes and to increase the power of the analysis by combining 2 outcomes of common etiology (15).

The objective of this study was to identify baseline participant characteristics predictive of 3 different definitions of symptomatic worsening following 26 weeks of participation in a nonsurgical chronic disease management program for hip and knee OA and to examine the consistency of predictors for the definitions. Previous studies have identified age, body mass index (BMI), and pain intensity as predictors of radiographic progression and TJA of knee (6–9) and hip OA (9,10). As such, these were pragmatically chosen as potential predictors of worsening following participation in the Osteoarthritis Chronic Care Program (OACCP). TJA waitlist status was also selected based on evidence that some patients report deterioration in health status following a ≥6 months wait for TJA (16). Additional predictors included signal joint, sex, number of comorbidities, a functional performance measure (6-Minute Walk Test [6MWT]), and presence of depression, which have all been previously associated with response to nonsurgical rehabilitation programs for hip and knee OA (17–19). We hypothesized that the same participants with similar demographic, psychological, disease-related, and functional performance predictor variables would be identified as "worse" across 3 definitions of worsening.

PATIENTS AND METHODS

Participants and data collection. This study comprised a cohort of consecutive participants with symptomatic and radiographic hip and knee OA recruited for the OACCP at Royal North Shore Hospital (RNSH). Participants were recruited directly from RNSH and Ryde Hospital (New South Wales, Australia) TJA waitlists or referral by rheumatologists, orthopedic surgeons, and general practitioners. People with a diagnosis of knee or hip OA were eligible if they reported pain in the affected knee/hip on most days of the past month (18). Data were included from participants who had completed at least 140 days in the program, and there were no exclusion criteria. Ethics approval for analysis of OACCP clinical data was provided by the New South Wales Population and Health Services Research Ethics Committee (AUREI reference HREC/12/CIPHS/63, Cancer Institute NSW, reference 2012/08/413).

The objectives of the OACCP were to reduce pain and increase function and quality of life of participants through the provision of tailored interventions delivered by a multidisciplinary team, including a physical therapist, rheumatologist, dietitian, occupational therapist, social worker, and orthotist. At the initial assessment, an experienced musculoskeletal physical therapist (MJW) provided participants with education about their OA and associated comorbidities, set patient-oriented goals, and prescribed behavioral modification strategies and an exercise program. The exercise program was comprised of strength and cardiovascular training and was progressed at 12-, 26-, and 52-week reassessments. Participants then attended a multidisciplinary clinic for consultation with a rheumatologist and a selection of other health professionals according to their individual clinical needs.

Outcome measures. Demographic data were recorded at baseline. The signal joint, i.e., the predominant site of OA, was determined by clinical and radiographic examination. Anthropometric measures were performed using a standardized protocol (20), including height, weight, waist and hip circumferences, and BMI. Participants rated their average pain on the day of assessment using a 10-cm visual analog scale (VAS; where 0 indicated no pain and 10 indicated the most pain imaginable) (21).

The validated, disease-specific Hip Disability and Osteoarthritis Outcome Score (22) and the Knee Injury and Osteoarthritis Outcome Score (23) require participants to rate their symptoms, stiffness, pain, physical function, recreational activities, and quality of life on 5-point Likert

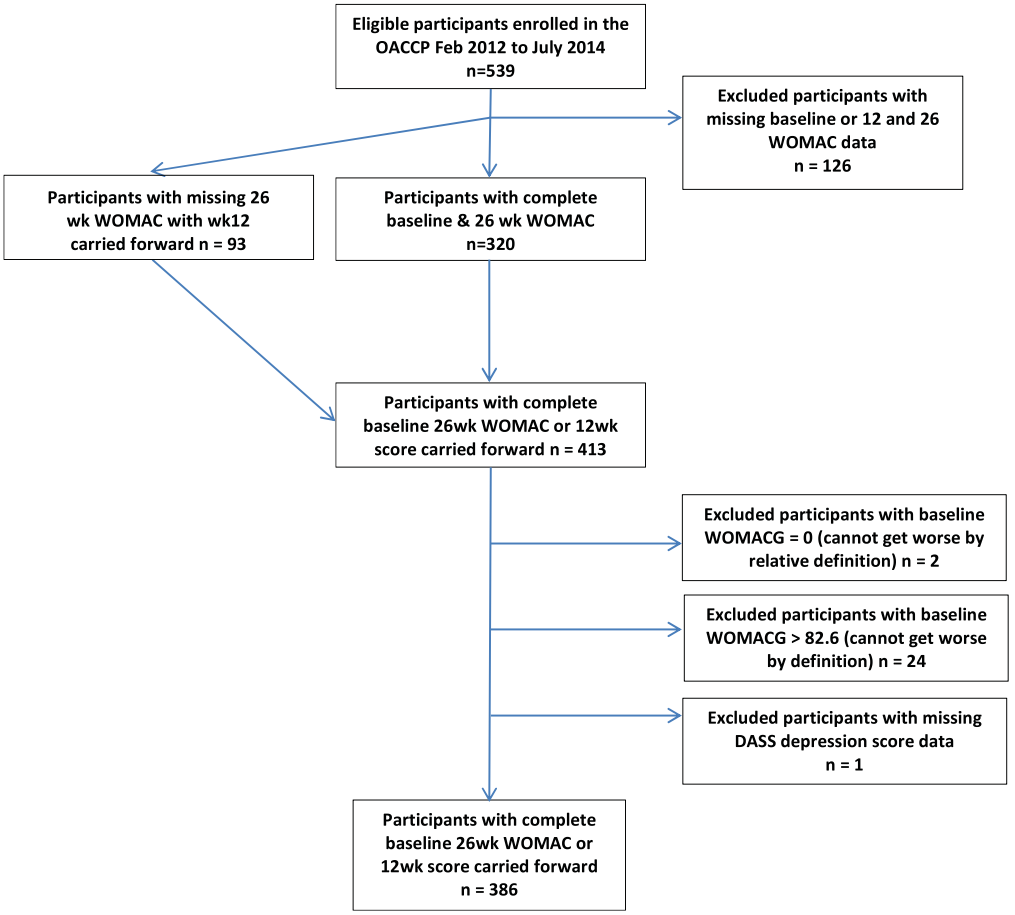


Figure 1. Study flow diagram. OACCP = Osteoarthritis Chronic Care Program; WOMACG = Western Ontario and McMaster Universities Osteoarthritis Index global score; DASS = Depression, Anxiety and Stress Scale.

scales. These questionnaires subsume the WOMAC questions, enabling conversion as follows:

$$\text{WOMAC-G} = (\text{sum of pain} + \text{stiffness} + \text{function items}) \times 100/96 \text{ (24,25)}$$

The 21-item version of the Depression, Anxiety and Stress Scale, using a 4-point Likert scale, asks participants to rate how much 21 separate statements applied to them over the past week. Subscores indicate the presence/absence of symptoms of depression, anxiety, and stress (26). Participants were dichotomized into those with depression subscores of 0–9 with no depressive symptoms, and those with subscores ≥10 with signs of depression (26).

A modified version of the Self-Administered Comorbidity Questionnaire (27) quantified the number of comorbidities experienced by each participant. Participants were asked to respond yes or no to the question, "Has your doctor told you that you have any of the following problems?" followed by a list of 21 commonly reported conditions, plus an "other" category. We categorized the number of comorbidities into groups of low (0–1), moderate (2–3), and high (≥4).

The 6MWT is recommended to assess long-distance walking and submaximal aerobic capacity (28). Participants were asked to walk as quickly as possible for 6 minutes on a flat 25-meter track with no corners (29), with the distance walked recorded in meters. Measurement of oxygen saturation, heart rate, and perceived exertion were taken prior to and at test completion. Participants with respiratory or cardiac comorbidity had measures at 1-minute intervals; the test was discontinued if participants reported concerning symptoms.

At reassessments, participants were asked a transition question ("Compared with when I started this program, my hip/knee has. . .") to rate their signal joint health status compared with prior to starting the program, with the following choices on a 7-point scale: much improved, moderately improved, slightly improved, not changed, slightly worse, moderately worse, or much worse.

Definitions of worsening. *WOMAC-G definition of worsening.* The minimal important difference (MID) is the smallest difference in scores of the variable considered to be beneficial or detrimental by participants (30). MID thresholds for worsening of 9.6 points absolute and 21% relative change in WOMAC-G were previously determined for a comprehensive rehabilitation intervention for hip and knee OA (13). We termed this threshold the "WOMAC-G

definition of worsening." A similar threshold has been used previously to indicate meaningful symptomatic worsening of participants on TJA waitlists (16).

Transition definition of worsening. In addition to pre- and postintervention assessment questionnaires, a transition scale is recommended to assess efficacy of self-management education programs (14). Therefore, the second definition of worsening was defined by the transition scale participant response of "moderately worse" or "much worse." In the absence of evidence for an ideal cutoff for change on the transition question that is meaningful to participants (31), we decided that "slightly worse" was not an adequate threshold for worsening, in an attempt to ensure that participants were reporting a change that was important to them.

Composite definition of worsening. The final definition of worsening was based on combined criteria of 9.6 points absolute and 21% relative change in WOMAC-G scores or "moderately worse" or "much worse" on the transition scale. This was chosen to reflect the dual importance of self-reported worsening of symptoms and self-reported overall deterioration of the signal joint.

Statistical analysis. Power calculations were based on evidence that 25% of patients waitlisted for TJA worsened ≥9.6 WOMAC-G points over 6 months (16). Given that only half of the OACCP participants were on TJA waitlists, we extrapolated that around 12.5% of the sample would report worsening. Further, because all participants received interventions for their OA, we expected that the majority of participants would report no change or improvements following the intervention and therefore estimated 10% of participants would "worsen." A sample of 500 participants was considered sufficient to include 3–5 variables in the final model, assuming 10 participants reported worsening per predictor variable (32,33).

A series of regression analyses were conducted in SPSS (version 22.0). For each model, the dependent variable was "worsening" and was based on dichotomization of participants into 2 groups: worse and not worse using each definition of worsening. Independent predictor variables were identified a priori, and the physical therapist collecting data was blinded to which variables were analyzed as predictors of worsening. Missing 26-week assessment data from patients who were not lost to followup had their 12-week assessment data carried forward, similar to intent-to-treat analysis. Further, the 6MWT results were standardized to ensure the scale (meters) was comparable with the other variables (34).

Univariate logistic regression analyses examined the odds of an OACCP participant worsening when each predictor variable was present. Subsequently, multivariate regression models were built for each definition of worsening. Variables exhibiting odds with P less than 0.2 that trended in the same direction across a minimum of 2 definitions were included in the base model for all 3 definitions of worsening. This method enabled interpretation and comparison of results between the composite definition and the single-outcome definitions (WOMAC-G and transition definitions) (15). The least significant predictor was removed at each step of the modelling until only significant variables remained. The regression coefficients of the remaining variables were checked on removal of each variable from the model, and in the presence of a change of ≥10% the variable was retained. Testing for interactions was performed by combining variables of interest. The validity of each model was assessed using Hosmer-Lemeshow goodness-of-fit, and the predictive ability of the model was assessed by calculation of sensitivity and specificity.

RESULTS

Of 539 participants consecutively enrolled in the OACCP from March 2012 to July 2014, 153 were excluded due to missing data or the presence of a floor/ceiling effect whereby they could not achieve the WOMAC-G definition (Figure 1). Reasons for missing data included undergoing TJA, medical advice, moving interstate, and illness or death.

Included and excluded participants were of similar age, were overweight, and reported a low to moderate number (0–3) of comorbidities (Table 1). Higher proportions of excluded participants were on TJA waitlists, presented with hips as the signal joint, and demonstrated depressive symptoms. Fewer excluded participants finished secondary school ($P = 0.018$) and their mean WOMAC-G scores were higher ($P < 0.001$). Excluded patients exhibited worse 6MWT results and higher baseline pain-VAS, although the mean difference was within measurement error for both outcomes (35). In the multidisciplinary OACCP clinic, most participants (95%) saw a rheumatologist, 75% were referred to a dietitian, 55% to an occupational therapist, 50% to a social worker, and 50% to an orthotist.

Definition of worsening outcomes. The 386 participants with complete data sets were included in the regression analyses. Of these, 34 (9%) reported worsening according to the WOMAC-G definition, 34 (9%) met the criteria for worsening for the transition definition, and 56 (15%) for the composite definition. Only 12 participants met all 3 definitions of worsening. According to both the transition definition (odds ratio [OR] 2.7 [95% confidence interval (95% CI) 1.26, 5.62]) and the composite definition (OR 2.2 [95% CI 1.22, 3.91]), OACCP participants waitlisted for TJA had more than twice the odds of reporting worsening. There was a trend that those with signal joint knees had lower odds of reporting worsening (Table 2). The standardized 6MWT was significantly associated with the transition definition and trended towards significance for the composite definition of worsening whereby farther walking distance indicated reduced odds of worsening (OR 0.7 [95% CI 0.46, 0.70]). The results of the univariate analyses differed when the WOMAC-G worsening definition was applied; participants reporting depressive symptoms were less likely to worsen.

Table 1. Baseline characteristics of OACCP participants*

Baseline characteristics	Included (n = 386)	Excluded: missing data (n = 153)†	*P*‡
Female	69	63	0.242
Age, mean ± SD years	66.3 ± 9.97	65.8 ± 11.35	0.620
Signal joint knee	85	73	0.006
On TJA waitlist	46	59	0.008
Residence			
Lives alone§	28	29	0.694
Speaks English¶	93	89	0.146
Engaged in paid employment#	31	28	0.573
Education			0.018
Finished secondary school or higher**	57	46	
Did not finish secondary school††	43	54	
BMI, mean ± SD kg/m^2	30.0 ± 6.47	32.1 ± 7.17	0.001
Pain VAS (range 0–10), mean ± SD	4.1 ± 2.17	5.4 ± 2.58	< 0.001
6MWT, mean ± SD meters	421.6 ± 111.72	380.7 ± 129.13	0.001
Depressive symptoms: DASS depression subscale (≥14) (%)	35	48	0.005
Number of comorbidities			0.233
Low (0–1)	39	32	
Moderate (2–3)	39	41	
High (≥4)	22	27	
Baseline WOMAC-G, mean ± SD (range)‡‡	49.1 ± 18.09 (3.1–82.3)	61.7 ± 23.67 (0–97.9)	< 0.001
Baseline WOMAC-G for those who worsened vs. didn't worsen, mean ± SD§§			
WOMAC-G			
Worse, mean ± SD (n = 34)	38.4 ± 16.34		
Not worse, mean ± SD (n = 352)	50.2 ± 17.93		
(CI difference in means) *P*	(−18.02, −5.45) < 0.01		
Transition¶¶			
Worse, mean ± SD (n = 34)	58.3 ± 17.31		
Not worse, mean ± SD (n = 352)	48.3 ± 17.94		
(CI difference in means) *P*	(3.69, 16.32) 0.002		
Composite##			
Worse, mean ± SD (n = 56)	49.3 ± 19.91		
Not worse, mean ± SD (n = 330)	49.1 ± 17.79		
(CI difference in means) *P*	(−4.91, 5.38) 0.929		

* Values are the percentage unless indicated otherwise. OACCP = Osteoarthritis Chronic Care Program; TJA = joint replacement arthroplasty; BMI = body mass index; VAS = visual analog scale; 6MWT = 6-Minute Walk Test; DASS = Depression, Anxiety and Stress Scale; WOMAC-G = Western Ontario and McMaster Universities Osteoarthritis Index global scores; CI = confidence interval.
† Missing data include those participants with missing 0-, 12-, or 26-week WOMAC-G data (n = 126), those whose week 0 WOMAC-G score equaled 0 (n = 2), those with week 0 WOMAC-G >82.6, and missing DASS depression data (n = 1).
‡ Independent analysis of variance or chi-square statistic comparing included and excluded participants.
§ Lives alone reported by participants. Living with others included living with abled/disabled person, in hostel or aged-care residential facility.
¶ Participants who did not speak English (~7%) required the use of an interpreter.
# Currently employed includes participants who reported engaging in full-/part-time paid work.
** Included participants who reported finishing secondary school (final year) or university degree.
†† Includes participants who did not finish secondary school and those who reported no formal schooling.
‡‡ WOMAC-G scores are calculated from the Hip Disability and Osteoarthritis Outcome Score and the Knee Injury and Osteoarthritis Outcome Score, where 0 indicates no problems and 100 indicates extreme problems.
§§ WOMAC-G minimum clinically important difference for worsening is 9.6 points absolute and 21% relative change in WOMAC-G compared to baseline.
¶¶ Transition definition of worsening; participant response was "moderately worse" or "much worse."
Composite definition of worsening; 9.6 points absolute and 21% relative change in WOMAC-G scores compared to baseline OR transition question response as "moderately worse" or "much worse."

Multivariate models. The final model for the WOMAC-G definition retained only presence of depressive symptoms (OR 0.30 [95% CI 0.11, 0.79]) (Table 3). The final multivariate model for the transition definition contained TJA waitlist, 6MWT, and signal joint; however, these predictors did not attain statistical significance (Table 3). The final multivariate composite definition model retained TJA waitlist (OR 1.91 [95% CI 1.04, 3.51]) and 6MWT (OR 0.83 [95% CI 0.60, 1.13]). According to the composite definition, participants on TJA waitlists had almost twice the odds of worsening. The 6MWT was a confounder, although as the CI crossed zero, we can interpret this as having little or no effect. The composite definition model fit the data well using the Hosmer-

Table 2. Univariate logistic regression analyses of predictors for 3 definitions of worsening*

Independent variable	WOMAC-G† Unadjusted OR (95% CI)	WOMAC-G† *P*	Transition‡ Unadjusted OR (95% CI)	Transition‡ *P*	Composite§ Unadjusted OR (95% CI)	Composite§ *P*
Age	1.0 (0.99, 1.08)	0.062	1.0 (0.99, 1.10)	0.114	1.0 (1.00, 1.07)¶	0.028¶
Sex						
Male	Reference		Reference		Reference	
Female	1.1 (0.50, 0.23)	0.895	1.6 (0.78, 3.29)	0.199	1.3 (0.70, 2.28)	0.477
Signal joint						
Hip	Reference		Reference		Reference	
Knee	1.0 (0.39, 2.84)	0.922	0.5 (0.20, 1.04)	0.063	0.5 (0.27, 1.07)	0.078
Pain VAS	1.0 (0.83, 1.15)	0.820	1.2 (1.04, 1.45)¶	0.017¶	1.1 (1.00, 1.24)	0.203
Number of comorbidities						
Low	Reference	0.274	Reference	0.605	Reference	0.365
Moderate	1.3 (0.63, 2.88)	0.443	0.9 (0.43, 1.1)	0.845	1.1 (0.60, 2.05)	0.753
High	0.5 (0.17, 1.73)	0.302	0.6 (0.21, 1.68)	0.323	0.6 (0.26, 1.41)	0.244
Depression						
No depression	Reference		Reference		Reference	
Any depression	0.3 (0.11, 0.79)¶	0.015¶	1.4 (0.66, 2.77)	0.409	0.6 (0.35, 1.22)	0.180
6MWT	1.1 (0.72, 1.50)	0.791	0.7 (0.46, 0.70)¶	0.033¶	0.8 (0.57, 1.03)	0.079
BMI	1.0 (0.95, 1.06)	0.989	1.0 (0.91, 1.02)	0.226	1.0 (0.93, 1.02)	0.351
TJA waitlist						
On list	Reference		Reference		Reference	
Not on list	1.4 (0.67, 2.73)	0.404	2.7 (1.26, 5.62)¶	0.011¶	2.2 (1.22, 3.91)¶	0.009¶

* WOMAC-G = Western Ontario and McMaster Universities Osteoarthritis Index global score; OR = odds ratio; 95% CI = 95% confidence interval; Reference = reference category for logistic regression; VAS = visual analog scale; 6MWT = standardized 6-Minute Walk Test; BMI = body mass index; TJA = total joint arthroplasty.
† WOMAC-G minimum clinically important difference for worsening: 9.6 points absolute and 21% relative change in WOMAC-G compared to baseline.
‡ Transition question definition of worsening: participant response as "moderately worse" or "much worse."
§ Composite definition of worsening: 9.6 points absolute and 21% relative change in WOMAC-G compared to baseline OR transition question response as "moderately worse" or "much worse."
¶ $P < 0.2$, one of the criteria for entry into multivariate model.

Lemeshow goodness-of-fit test ($\chi^2 = 8.68$, 2 df, $P = 0.37$); however, it was only capable of explaining 5% of the variance in demonstrating worsening (Figure 2). No model could predict worsening on all 3 definitions together. Although the best model was specific, i.e., correctly identified participants who did not worsen, it lacked sensitivity, so failed to identify those who had worsened.

Table 3. Final multivariate logistic regression models for 3 definitions of worsening*

Definition of worse	Worse, no. (%)	Predictors	β coefficient	*P*	Adjusted OR	95% CI
WOMAC-G†	34 (9)	Constant	−2.04			
		Depression	−1.21	0.015	0.30	0.11, 0.79
Transition‡	34 (9)	Constant	−1.14			
		TJA waitlist	0.65	0.114	1.91	0.86, 4.29
		Signal joint knee	−0.54	0.238	0.58	0.24, 1.43
		6MWT	−0.34	0.092	0.071	0.48, 1.06
Composite§	56 (14.5)	Constant	−1.47			
		TJA waitlist	0.65	0.036	1.91	1.04, 3.51
		6MWT	−0.19	0.231	0.83	0.60, 1.13

* OR = odds ratio; 95% CI = 95% confidence interval; WOMAC-G = Western Ontario and McMaster Universities Osteoarthritis Index global score; TJA = total joint arthroplasty; 6MWT = 6-Minute Walk Test.
† WOMAC-G minimum clinically important difference for worsening: 9.6 points absolute and 21% relative change in WOMAC-G compared to baseline.
‡ Transition question definition of worsening: participant response was "moderately worse" or "much worse."
§ Composite definition of worsening: 9.6 points absolute and 21% relative change in WOMAC-G compared to baseline OR transition question response as "moderately worse" or "much worse."

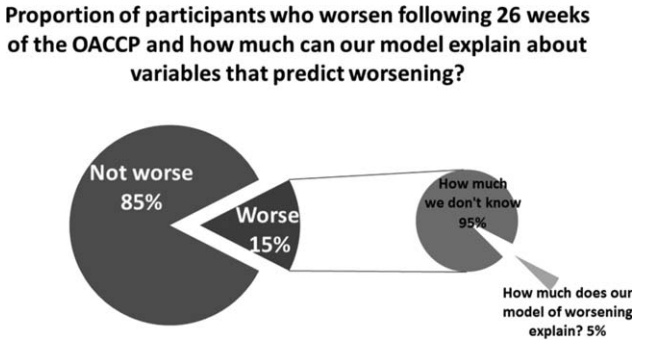


Figure 2. Summary of how much we can explain worsening in response to a nonsurgical program. OACCP = Osteoarthritis Chronic Care Program.

DISCUSSION

This study aimed to identify participant characteristics predictive of 3 definitions of worsening following 26 weeks of participation in the OACCP. We hypothesized that the same participants with similar demographic, psychological, disease-related, and functional performance predictor variables would be identified as "worse" across 3 definitions; however, this was not the case. Similar trends were demonstrated by the transition and composite models; TJA waitlist and 6MWT were retained in both, though the transition model also retained signal joint and did not reach statistical significance. These trends were not apparent when the WOMAC-G definition was applied.

The evidence for symptomatic deterioration while waiting for TJA is conflicting. A systematic review (36) reported people with hip or knee OA waiting for less than 6 months did not experience 10% deterioration in WOMAC pain or function scores. In contrast, 25% of those waiting for TJA longer than 6 months reported a decline of ≥9.6% WOMAC-G scores (16). Only 46% of our sample was listed for TJA and these participants had been waiting for approximately 6 months. This may account for the lack of association between TJA waitlist status and the WOMAC-G definition of worsening. The participants lost to followup due to TJA and other reasons reported a higher mean ± SD WOMAC-G of 61.7 ± 23.67, which is more similar to the previous study (16). These participants may have worsened in this time without surgery; however, we do not have the data to support this supposition. The group reporting worsening according to the WOMAC-G definition had significantly lower baseline mean ± SD WOMAC-G scores of 38.4 ± 16.34, compared to 50.2 ± 17.93 for those who did not (Table 1). One possible explanation is that there was a regression to the mean in participants with lower baseline WOMAC scores.

In contrast to our hypothesis, 22 of 34 people who reported worsening on the transition scale were not considered worse according to the WOMAC-G definition despite using "moderately worse" as the minimum cutoff. This contrasts with the study from which the threshold is derived (13). In the former study participants were not on the TJA waitlist, the intervention focused on physical therapy (not multidisciplinary), and was of shorter duration (3–4 weeks). It is possible that the thresholds we used to indicate worsening were not ideal for our study population. Further work is required to confirm the most appropriate threshold of worsening for this population.

Alternatively it is possible that discordance of worsening according to WOMAC-G and transition scale definitions may be attributed to the attitudes and expectations of participants waitlisted for TJA. The act of booking a person for TJA possibly preconditions them to believe that their signal joint should become worse over time. The language used, such as "end-stage," "severe," and "bone-on-bone," may influence their response of "moderately worse" or "much worse," even though their WOMAC-G scores, a much lengthier questionnaire directly asking about specific symptoms, did not reflect this. The association between use of specific language by caregivers and patient perception of disease severity has received minimal research attention and is a potential area for future work. However, considerable paternalism persists in medical decision-making about TJA (37,38). Perhaps some of our cohort believed that they needed the surgery because the surgeon said so, and this was sufficient evidence to report that their overall hip or knee joint health was "worse."

Participants on TJA waitlists had twice the odds of meeting the composite criteria for "worsening" following participation in the OACCP. While TJA can provide good symptomatic relief for most people with end-stage OA (39), nonsurgical management is efficacious in reducing the signs and symptoms of knee OA (5). It is advocated that patients be referred for nonsurgical treatments as a first port-of-call, and participation in chronic disease management programs should commence earlier in the OA disease course prior to being waitlisted for TJA. Although at this stage we can only explain a very small proportion of variables that predict worsening, referring participants earlier in their disease course may lessen their odds of worsening despite taking part in self-management programs.

The absence of depression was a significant predictor of worsening according to the WOMAC-G definition. This finding is counterintuitive; a possible explanation is that of 34 participants who worsened according to the WOMAC-G definition, only 5 reported depressive symptoms. This result is likely to be a type I error. The evidence for the absence of depression as a predictor of positive response to nonsurgical interventions for hip and knee OA is conflicting (18,19,40). Further investigation of the relationship between symptoms of depression and outcomes following participation in the OACCP is warranted. Depression has been associated with noncompliance with treatment in populations with chronic disease (41). Compliance with OACCP interventions was not measured in this study. Exploration of the association between compliance with multimodal therapies, depression, and self-reported worsening should also be addressed in future research.

Signal joint was independently associated with the transition and composite definitions of worsening; participants with knee OA had lower odds of worsening following the OACCP. However signal joint was not a significant predictor in the multivariate models. Previous research found that people waiting for hip TJA do not deteriorate compared to those awaiting knee TJA; however, those participants were not receiving interventions for their OA (42). In contrast, a

previous study following 26 weeks of participation in the OACCP for those with signal joint knees had twice the odds of responding as those with signal joint hips (18). Although exercise for hip OA may confer some reduction in pain and improvement in function, the treatment effect sizes are small (43). Further research into effective nonsurgical management options for participants with hip OA is urgently required.

Although previous research reports age, sex, and BMI to be important characteristics in disease progression and response to intervention for hip and knee OA (6,10,18,19,44,45), they were not found to be significant predictors of worsening in our current investigation. The presence of comorbidities has been associated with poorer health-related quality of life for OA patients (46), yet greater number of comorbidities was not associated with any definition of worsening. This is an important finding suggesting that nonsurgical management may be considered for anyone with any number of concomitant conditions. Increasing pain over time has been associated with progression to TJA (7), and baseline WOMAC pain scores independently correlated with TJA 6 years following assessment (9). Baseline pain-VAS was only independently associated with the transition definition, and therefore failed to meet the criteria for entry into the multivariate models. The 6MWT was not a predictor of "response" in a similar OACCP cohort (47); however, the self-paced 40-meter walk test was predictive of response to physical therapy interventions for patients with hip OA (17). There is a gap in the research concerning external validity of the 6MWT as compared to PROs (48).

This study has several notable strengths: overall it was well-powered, potential predictor variables were identified a priori, and care was taken to avoid overfitting the prediction models. A physician's diagnosis of OA was used, which has good face validity (49). Although we were able to determine why participants withdrew participation from the OACCP, WOMAC data were missing for some who progressed to TJA, resulting in their exclusion from the analysis. This potentially limits the applicability of our results. Data from these participants would have been very useful to determine if these participants met any of the "worsening" definitions immediately prior to their surgery. Significant predictors of worsening were found; however, the multivariate models provide a very small proportion of the factors that predict worsening (Figure 2) and were not sensitive.

It is possible that stronger predictors of worsening exist that have not yet been studied, or perhaps the thresholds chosen to represent the predictors affected the outcome. We transformed symptoms of depression and number of comorbidities into categorical variables, which may not have been ideal. This study did not include a control group and was limited to 1 OACCP site; both investigation into predictors of worsening compared to a control group and use of more heterogeneous samples of participants are important areas for future research.

Three definitions of worsening were applied; potential predictors were identified only when using the composite definition of worsening. While TJA waitlist status was associated with a 2-fold increase in odds of reporting worsening using this definition, the model explained only 5% of the total variance. Further, following 26 weeks of participation in the OACCP, the WOMAC-G was largely discordant with the transition and composite definitions of worsening. Participants with similar demographic, psychological, disease-related, and functional performance predictor variables were not consistently identified as "worse" across the 3 definitions. Variables that predict worsening are largely unknown and further research into this area is warranted in order to present comprehensive and targeted management of patients with hip and knee OA.

ACKNOWLEDGMENT

Data for this analysis were obtained with permission of the Agency for Clinical Innovation (ACI). The authors wish to thank the ACI Musculoskeletal Network, especially Mary Fien and Robyn Speerin for their roles in development and oversight of the database and for organizing access to the data.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be submitted for publication. Ms Eyles had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Eyles, Mills.

Acquisition of data. Eyles, Williams, Hunter.

Analysis and interpretation of data. Eyles, Mills, Lucas, Makovey, Teoh, Hunter.

REFERENCES

1. World Health Organization. Chronic diseases and health promotion: chronic rheumatic conditions. URL: http://www.who.int/chp/topics/rheumatic/en/.
2. Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2013;380:2163–96.
3. Hochberg MC, Altman RD, April KT, Benkhalti M, Guyatt G, McGowan J, et al. American College of Rheumatology 2012 recommendations for the use of nonpharmacologic and pharmacologic therapies in osteoarthritis of the hand, hip, and knee. Arthritis Care Res (Hoboken) 2012;64:465–74.
4. Larmer PJ, Reay ND, Aubert ER, Kersten P. Systematic review of guidelines for the physical management of osteoarthritis. Arch Phys Med Rehabil 2014;95:375–89.
5. Zhang W, Moskowitz RW, Nuki G, Abramson S, Altman RD, Arden N, et al. OARSI recommendations for the management of hip and knee osteoarthritis, part I: critical appraisal of existing treatment guidelines and systematic review of current research evidence. Osteoarthritis Cartilage 2007;15:981–1000.
6. Chapple CM, Nicholson H, Baxter GD, Abbott JH. Patient characteristics that predict progression of knee osteoarthritis: a systematic review of prognostic studies. Arthritis Care Res (Hoboken) 2011;63:1115–25.
7. Raynauld JP, Martel-Pelletier J, Haraoui B, Choquette D, Dorais M, Wildi LM, et al. Risk factors predictive of joint replacement in a 2-year multicentre clinical trial in knee osteoarthritis using MRI: results from over 6 years of observation. Ann Rheum Dis 2011;70:1382–8.
8. Conaghan PG, D'Agostino MA, le Bars M, Baron G, Schmidely N, Wakefield R, et al. Clinical and ultrasonographic

predictors of joint replacement for knee osteoarthritis: results from a large, 3-year, prospective EULAR study. Ann Rheum Dis 2009;69:644–7.
9. Hawker GA, Guan J, Croxford R, Coyte PC, Glazier RH, Harvey BJ, et al. A prospective population-based study of the predictors of undergoing total joint arthroplasty. Arthritis Rheum 2006;54:3212–20.
10. Wright AA, Cook C, Abbott JH. Variables associated with the progression of hip osteoarthritis: a systematic review. Arthritis Care Res (Hoboken) 2009;61:925–36.
11. Gossec L, Paternotte S, Bingham CO III, Clegg DO, Coste P, Conaghan PG, et al. OARSI/OMERACT initiative to define states of severity and indication for joint replacement in hip and knee osteoarthritis: an OMERACT 10 special interest group. J Rheumatol 2011;38:1765–9.
12. Manno RL, Bingham CO III, Paternotte S, Gossec L, Halhol H, Giacovelli G, et al. OARSI-OMERACT initiative: defining thresholds for symptomatic severity and structural changes in disease modifying osteoarthritis drug (DMOAD) clinical trials. Osteoarthritis Cartilage 2012;20:93–101.
13. Angst F, Aeschlimann A, Michel BA, Stucki G. Minimal clinically important rehabilitation effects in patients with osteoarthritis of the lower extremities. J Rheumatol 2002;29: 131–8.
14. Jordan JE, Haynes K, Livingston JA, Osborne RH. Comparison of the pre-post and transition question assessments in a health education setting. J Clin Epidemiol 2009;62:642–9.
15. Friedman L, Furberg C, DeMets D. Fundamentals of clinical trials, 4th ed. New York: Springer; 2010.
16. Ackerman IN, Bennell KL, Osborne RH. Decline in health-related quality of life reported by more than half of those waiting for joint replacement surgery: a prospective cohort study. BMC Musculoskelet Disord 2011;12:108.
17. Wright AA, Cook CE, Flynn TW, Baxter GD, Abbott JH. Predictors of response to physical therapy intervention in patients with primary hip osteoarthritis. Phys Ther 2011;91: 510–24.
18. Eyles JP, Lucas BR, Patterson JA, Williams MJ, Weeks K, Fransen M, et al. Does clinical presentation predict response to a nonsurgical chronic disease management program for endstage hip and knee osteoarthritis? J Rheumatol 2014;41: 2223–31.
19. Weigl M, Angst F, Aeschlimann A, Lehmann S, Stucki G. Predictors for response to rehabilitation in patients with hip or knee osteoarthritis: a comparison of logistic regression models with three different definitions of responder. Osteoarthritis Cartilage 2006;14:641–51.
20. New South Wales Agency for Clinical Innovation. Osteoarthritis chronic care program pilot site manual, version 8. 2012.
21. Hawker GA, Mian S, Kendzerska T, French M. Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP). Arthritis Care Res (Hoboken) 2011;63 Suppl 11:S240–52.
22. Thorborg K, Roos EM, Bartels EM, Petersen J, Holmich P. Validity, reliability and responsiveness of patient-reported outcome questionnaires when assessing hip and groin disability: a systematic review. Br J Sports Med 2010;44: 1186–96.
23. Roos EM, Lohmander LS. The Knee injury and Osteoarthritis Outcome Score (KOOS): from joint injury to osteoarthritis. Health Qual Life Outcomes 2003;1:64.
24. KOOS user's guide. 2003. URL: www.koos.nu/KOOS-Guide2003.pdf.
25. HOOS user's guide. 2003. URL: www.koos.nu/HOOSScoring-2013.pdf.
26. Lovibond PF, Lovibond SH. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. Behav Res Ther 1995;33:335–43.
27. Sangha O, Stucki G, Liang MH, Fossel AH, Katz JN. The Self-Administered Comorbidity Questionnaire: a new method to assess comorbidity for clinical and health services research. Arthritis Rheum 2003;49:156–63.
28. Dobson F, Hinman RS, Roos EM, Abbott JH, Stratford P, Davis AM, et al. OARSI recommended performance-based tests to assess physical function in people diagnosed with hip or knee osteoarthritis. Osteoarthritis Cartilage 2013;21: 1042–52.
29. Bennell K, Dobson F, Hinman R. Measures of physical performance assessments: Self-Paced Walk Test (SPWT), Stair Climb Test (SCT), Six-Minute Walk Test (6MWT), Chair Stand Test (CST), Timed Up & Go (TUG), Sock Test, Lift and Carry Test (LCT), and Car Task. Arthritis Care Res (Hoboken) 2011;63 Suppl 11:S350–70.
30. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR, and the Clinical Significance Consensus Meeting Group. Methods to explain the clinical significance of health status measures. Mayo Clin Proc 2002;77:371–83.
31. Terwee CB, Roorda LD, Dekker J, Bierma-Zeinstra SM, Peat G, Jordan KP, et al. Mind the MIC: large variation among populations and methods. J Clin Epidemiol 2010;63:524–34.
32. Peduzzi P, Concato J, Feinstein AR, Holford TR. The importance of events per independent variable in multivariable analysis, II: accuracy and precision of regression estimates. J Clin Epidemiol 1995;48:1503–10.
33. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol 1996;49: 1373–9.
34. Tabachnick BG, Fidell LS. Using multivariate statistics, 5th ed. New York: Allyn & Bacon; 2006.
35. Naylor JM, Hayen A, Davidson E, Hackett D, Harris IA, Kamalasena G, et al. Minimal detectable change for mobility and patient-reported tools in people with osteoarthritis awaiting arthroplasty. BMC Musculoskelet Disord 2014;15: 235.
36. Hoogeboom TJ, van den Ende CH, van der Sluis G, Elings J, Dronkers JJ, Aiken AB, et al. The impact of waiting for total joint replacement on pain and functional status: a systematic review. Osteoarthritis Cartilage 2009;17:1420–7.
37. Fowler FJ Jr, Gerstein BS, Barry MJ. How patient centered are medical decisions? Results of a national survey. JAMA Intern Med 2013;173:1215–21.
38. Sepucha K, Feibelmann S, Chang Y, Clay CF, Kearing SA, Tomek I, et al. Factors associated with the quality of patients' surgical decisions for treatment of hip and knee osteoarthritis. J Am Coll Surg 2013;217:694–701.
39. Escobar A, Gonzalez M, Quintana JM, Vrotsou K, Bilbao A, Herrera-Espineira C, et al. Patient acceptable symptom state and OMERACT-OARSI set of responder criteria in joint replacement: identification of cut-off values. Osteoarthritis Cartilage 2012;20:87–92.
40. Wolf S, Foley S, Budiman-Mak E, Moritz T, O'Connell S, Jelinek C, et al. Predictors of weight loss in overweight veterans with knee osteoarthritis who participated in a clinical trial. J Rehabil Res Dev 2010;47:171–81.
41. Grenard JL, Munjas BA, Adams JL, Suttorp M, Maglione M, McGlynn EA, et al. Depression and medication adherence in the treatment of chronic diseases in the United States: a meta-analysis. J Gen Intern Med 2011;26:1175–82.
42. Kapstad H, Rustoen T, Hanestad BR, Moum T, Langeland N, Stavem K. Changes in pain, stiffness and physical function in patients with osteoarthritis waiting for hip or knee joint replacement surgery. Osteoarthritis Cartilage 2007;15: 837–43.
43. Fransen M, McConnell S, Hernandez-Molina G, Reichenbach S. Exercise for osteoarthritis of the hip. Cochrane Database Syst Rev 2014;4:CD007912.
44. Gandhi R, Wasserstein D, Razak F, Davey JR, Mahomed NN. BMI independently predicts younger age at hip and knee replacement. Obesity (Silver Spring) 2010;18:2362–6.
45. Riddle DL, Stratford PW. Body weight changes and corresponding changes in pain and function in persons with

symptomatic knee osteoarthritis: a cohort study. Arthritis Care Res (Hoboken) 2013;65:15–22.
46. Tuominen U, Blom M, Hirvonen J, Seitsalo S, Lehto M, Paavolainen P, et al. The effect of co-morbidities on health-related quality of life in patients placed on the waiting list for total joint replacement. Health Qual Life Outcomes 2007;5:16.
47. Eyles J, Lucas B, Hunter DJ. Targeting care tailoring nonsurgical management according to clinical presentation. Rheum Dis Clin North Am 2013;39:213–33.
48. Dobson F, Hinman RS, Hall M, Terwee CB, Roos EM, Bennell KL. Measurement properties of performance-based measures to assess physical function in hip and knee osteoarthritis: a systematic review. Osteoarthritis Cartilage 2012; 20:1548–62.
49. Jordan K, Clarke AM, Symmons DP, Fleming D, Porcheret M, Kadam UT, et al. Measuring disease prevalence: a comparison of musculoskeletal disease using four general practice consultation databases. Br J Gen Pract 2007;57:7–14.

Appendix 6: Instruments assessing attitudes toward or capability regarding self-management of osteoarthritis: a systematic review of measurement properties.

Osteoarthritis and Cartilage 25 (2017) 1210–1222

Osteoarthritis and Cartilage

Review

Instruments assessing attitudes toward or capability regarding self-management of osteoarthritis: a systematic review of measurement properties

J.P. Eyles † ‡ § *, D.J. Hunter † ‡, S.R.F. Meneses † ‡, N.J. Collins ||, F. Dobson ¶, B.R. Lucas † §, K. Mills #

† *Kolling Institute of Medical Research, Institute of Bone and Joint Research, University of Sydney, Australia*
‡ *Department of Rheumatology, Royal North Shore Hospital and Northern Clinical School, University of Sydney, Australia*
§ *Physiotherapy Department, Royal North Shore Hospital, Sydney, Australia*
|| *School of Health & Rehabilitation Sciences, University of Queensland, Australia*
¶ *Centre for Health, Exercise and Sports Medicine, Department of Physiotherapy, School of Health Sciences, The University of Melbourne, Australia*
# *Centre for Physical Health, Department of Medicine and Health Sciences, Macquarie University, Sydney, Australia*

ARTICLE INFO

Article history:
Received 11 November 2016
Accepted 22 February 2017

Keywords:
Self-management
Instruments
Measurement properties
Psychometrics
Clinimetrics
Systematic review

SUMMARY

Objective: To make a recommendation on the "best" instrument to assess attitudes toward and/or capabilities regarding self-management of osteoarthritis (OA) based on available measurement property evidence.

Methods: Electronic searches were performed in MEDLINE, EMBASE, CINAHL and PsychINFO (inception to 27 December 2016). Two reviewers independently rated measurement properties using the Consensus-based Standards for the selection of Health Measurement Instruments (COSMIN) 4-point scale. Best evidence synthesis was determined by considering COSMIN ratings for measurement property results and the level of evidence available for each measurement property of each instrument.

Results: Eight studies out of 5653 publications met the inclusion criteria, with eight instruments identified for evaluation: Multidimensional Health Locus of Control (MHLC), Perceived Behavioural Control (PBC), Patient Activation Measure (PAM), Educational Needs Assessment (ENAT), Stages of Change Questionnaire in Osteoarthritis (SCQOA), Effective Consumer Scale (EC-17) and Perceived Efficacy in Patient–Physician Interactions five item (PEPPI-5) and ten item scales. Measurement properties assessed for these instruments included internal consistency ($k = 8$), structural validity ($k = 8$), test–retest reliability ($k = 2$), measurement error ($k = 1$), hypothesis testing ($k = 3$) and cross-cultural validity ($k = 3$). No information was available for content validity, responsiveness or minimal important change (MIC)/minimal important difference (MID). The Dutch PEPPI-5 demonstrated the best measurement property evidence; strong evidence for internal consistency and structural validity but limited evidence for reliability and construct validity.

Conclusion: Although PEPPI-5 was identified as having the best measurement properties, overall there is a poor level of evidence currently available concerning measurement properties of instruments to assess attitudes toward and/or capabilities regarding osteoarthritis self-management. Further well-designed studies investigating measurement properties of existing instruments are required.

Introduction

Healthcare systems currently face a rising number of people living with chronic conditions leading to disability, without causing death[1]. The Chronic Care Model (CCM) has been promoted to assist healthcare systems to meet the escalating demands attributable to chronic conditions[2]. The CCM describes healthcare whereby patients are enabled to manage their condition supported by a

* Address correspondence and reprint requests to: J. Eyles, 7C Clinical Administration, Rheumatology Dept, Royal North Shore Hospital, St Leonards, NSW 2065, Australia.
E-mail addresses: jillian.eyles@sydney.edu.au (J.P. Eyles), david.hunter@sydney.edu.au (D.J. Hunter), sarah.meneses@sydney.edu.au (S.R.F. Meneses), n.collins1@uq.edu.au (N.J. Collins), fdobson@unimelb.edu.au (F. Dobson), Barbara.Lucas@health.nsw.gov.au (B.R. Lucas), kathryn.mills@mq.edu.au (K. Mills).

http://dx.doi.org/10.1016/j.joca.2017.02.802
1063-4584/

proactive healthcare delivery system, involving a coordinated team of health professionals with the expertise required to provide decision support, all underpinned by appropriate health information systems[2]. Self-management programmes are interventions based on the tenets of the CCM; they aim to improve self-management capabilities. It follows that the efficacy of these programmes should be measured by assessing change in participants' attitudes toward and/or capabilities to manage their health. However, there are few recommendations guiding which instruments accurately measure self-management[3]. The widespread heterogeneity in standardised instruments measuring self-management programs is surprising given that the primary aim of these programs is to directly influence the attitudes toward and abilities to manage one's health.

This situation is apparent in self-management programmes for osteoarthritis (OA). Research examining the efficacy of OA self-management programmes has focussed on measures of pain and function[4]. While these outcomes are obviously important to this population, there appears to be disparity in the aims of self-management programmes and the outcomes used to assess efficacy[5]. Self-management programs aim to provide participants with the necessary tools to manage their own condition rather than "cure" OA. Although these programmes may not dramatically reduce pain and enhance functional ability, this does not necessarily reflect a failed strategy if the participants improve their attitudes towards and ability to manage symptoms and live with an acceptable quality of life despite their disease[5].

A systematic review reported low-to-moderate quality evidence of no or small benefits to participants of OA self-management education programmes[5]. The authors highlighted the heterogeneity of outcomes used to quantify the effects of self-management programmes and that work is needed to establish which outcomes are important to patients. This review recommended rigorous evaluation of OA self-management programmes with validated instruments fit to measure attitudes towards/capabilities to self-manage OA, and advised that to achieve this, the measurement properties of the existing instruments need further investigation[5].

Measurement properties refer to the ability of the instrument to truthfully and comprehensively measure the specified construct[6]. In addition, it is necessary to demonstrate that the instrument is discriminative, sensitive, reliable and deemed feasible in terms of cost and time constraints[7]. It is important to consider that the measurement properties of an instrument are not universal across different populations; hence, it cannot be assumed that one with good measurement properties in a specific population will demonstrate the same results in a different population[8]. Therefore, the measurement properties of an instrument must be considered within the specific context of the population of interest.

The aims of this systematic review were to: (1) identify studies reporting measurement properties of instruments assessing attitudes toward and/or capabilities regarding self-management of OA; (2) systematically critique the studies evaluating instruments using the Consensus-based Standards for the selection of Health Measurement Instruments (COSMIN) tool; and (3) synthesize the evidence available with the possibility of making rudimentary recommendations concerning the best evidence-based instruments to assess attitudes toward and/or capabilities regarding self-management of OA.

Methodology

Terminology

Self-management was defined as the individual's ability to manage their physical and psychological symptoms, treatments, consequences and lifestyle changes required to live with their OA[9]. Attitudes toward and/or capabilities regarding self-management of OA included the following constructs: knowledge, skills, beliefs, behaviours, activation, self-efficacy, health locus of control, readiness to change healthcare behaviours, healthcare navigation, participation, engagement, and motivation. This list of possible constructs was developed *a priori* using existing content knowledge about available instruments of the authors, and new constructs identified during the review were also included.

Review protocol

The review protocol was developed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement and prospectively registered with PROSPERO on 24 November 2015 (CRD42015019074).

Literature search

The review search strategy was developed and refined by the study authors according to the PRISMA statement and recommendations made for conducting systematic reviews of measurement properties[8,10]. Electronic searches were conducted of the following four bibliographic databases from inception to 27 December 2016: MEDLINE (PubMed), EMBASE (OvidSP), CINAHL (Ebsco), PsychINFO (OvidSP). An initial search was conducted using four main filters containing key search terms as briefly summarised below (see Appendix 1 PubMed search strategy):

I. **Construct** – attitudes toward and capabilities regarding self-management of OA using terms such as: "self-treatment OR self-management OR patient education…" Names of known instruments measuring attitudes and/or capabilities regarding self-management were added using 'OR': "health education impact questionnaire OR patient activation measure OR effective consumer scale …"
II. **Target population** – osteoarthritis OR osteoarth* OR degenerative arthritis OR arthrosis.
III. **Measurement instrument filter** – designed for PubMed to retrieve more than 97% of publications related to measurement properties[11] using terms such as: "instrumentation OR methods OR validation studies…" The filter was translated into the language of the other databases used.
IV. **Exclusion filter** – an exclusion filter was used to improve the precision of the measurement instrument filter[11].

Secondary searching was conducted for all instruments measuring attitudes toward and capabilities regarding self-management of OA identified during the initial search. The name of each instrument was used as the keyword combined (AND) with the target population filter in PubMed. Targeted hand searching of reference lists was also used. Results of the database searches were imported into Endnote X7 (Thomson Reuters, Philadelphia, USA).

Eligibility criteria

Study titles were screened by one reviewer (JE). Two reviewers (JE & SM) independently screened abstracts, followed by the full text of potentially eligible studies. Disagreements were discussed and resolved with a third reviewer (KM). Studies were included if they met the following criteria:

1. **Construct** – at least one instrument attempted to measure the participants' attitudes and/or capabilities regarding self-management of their OA.

2. **Target population** – adults diagnosed with any stage of OA according to American College of Rheumatology guidelines, clinical diagnosis of OA from examination findings, patients' symptoms or radiographic evidence of disease. Studies with mixed disease populations were excluded if the proportion of participants with a main diagnosis of OA was less than 80% and the results for OA participants were not reported separately.
3. **Measurement instrument** – patient-reported outcomes (PROs) (completed by the participant) in the form of questionnaires or scales.
4. **Measurement properties** – the study was required to explicitly state a primary or secondary aim to develop an instrument or examine at least one measurement property of the instrument involved.
5. **Setting** – the instrument was required to have been utilised in a clinic, field or community setting using readily available equipment. Instruments with a license fee were included.
6. **Publication type** – full text studies published as original articles in peer-reviewed journals.
7. **Language** – English language studies were included. Non-English language studies were noted and data extraction performed when possible, however these were excluded from COSMIN rating due to lack of access to translation resources, and the high level of detail required for a COSMIN review.

Data extraction

Two reviewers (JE & SM) independently extracted data to a predefined spreadsheet with a third reviewer (KM) available to resolve differences. The generalisability of the included studies was considered by extracting characteristics such as mean age, gender distribution, OA stage, setting and language. Relevant data regarding interpretability issues was extracted including distribution of scores, floor and ceiling effects, change scores, and minimal important change (MIC) or minimal important difference (MID)[12].

Methodological quality evaluation of the studies

Two raters (JE & NC) independently assessed the methodological quality of the included studies, with a third rater (FD) available to resolve discrepancies. Included studies were assessed according to the COSMIN taxonomy of the following measurement properties: internal consistency, reliability, measurement error, content validity, structural validity, hypotheses testing (a form of construct validity), cross-cultural validity, and responsiveness[13]. The definitions of these measurement properties are summarised in Table I[12]. Each measurement property featured within a particular study was rated separately according to the COSMIN tool; a robust quality evaluation tool using a 4-point scoring system: “poor”, “fair”, “good” or “excellent”[12,14]. An overall quality score was given for each measurement property in each study using the “worst score counts” method that accounted for the lowest rating of any item within that measurement property section[14].

Evaluation of measurement property result

An overall quality rating of the measurement property results for each instrument was performed using a checklist of criteria for good measurement properties[15] (Appendix 2). Two raters determined the quality rating using this additional tool (JE & SM) with disagreements resolved with a third reviewer (NC).

Data synthesis

Qualitative analysis

To summarise the level of evidence of each measurement property for each instrument, a “best evidence synthesis” was performed. The “best evidence synthesis” was derived by triangulating the methodological quality of the studies[12] (using the COSMIN score), the quality criteria for rating the results of measurement properties (Appendix 2)[15], and the level of evidence for the measurement properties of the instruments according to the following: “strong”, “moderate”, “limited”, “conflicting”, or “indeterminate”[8,15] (Table II).

Quantitative analysis

Meta-analysis of data was planned for studies of fair or better methodological quality and of sufficient homogeneity[8].

Results

The initial search strategy identified 5653 studies (Fig. 1). Following title and abstract screening, 44 studies were identified for full-text review. Following full-text review, eight studies were included[16–23]. Each study assessed a different instrument, therefore it was not possible to pool data for quantitative analyses.

The content of instruments varied widely with respect to the constructs of self-management they represented. Table III provides a content comparison of the constructs represented in the eight instruments, their characteristics are summarised in Table IV. The Patient Activation Measure (PAM)[16] required a license fee; all others were freely available online or following contact with the authors. Many instruments were translated into a language other

Table I
Definitions of measurement properties

Measurement property	Definition
Internal consistency	The degree to which items of an instrument are related to each other
Reliability	The proportion of the total variance of “true differences” measured by the instrument that is not attributed to measurement error
Measurement error	The component of a patient's score that is not due to real changes of the construct measured by the instrument, but attributed to systematic and/or random error
Content validity	The degree to which the content of the instrument measures the construct it intends to measure
Structural validity	The extent to which the scores of an instrument conform to the dimensionality of the construct intended
Hypotheses testing	An aspect of construct validity; when questions are formulated *a priori* about the expected relationships with instruments measuring related constructs
Cross-cultural validity	The extent to which the translated or culturally adapted instrument reflects the performance of the original version of the instrument
Criterion validity	When the scores of an instrument are compared to determine if they are reflective of the outcomes of another instrument considered to be the “gold standard”
Responsiveness	The measurement of the ability of the instrument to detect changes in scores that reflect change in the construct over time
Floor and ceiling effects	The proportion of participants who responded with the lowest or highest possible score on the instrument

Definitions adapted from Mokkink *et al.* J Clin Epidem 36 (2010) and de Vet, H. *et al.*, “*Measurement in Medicine: A Practical Guide to Biostatistics and Epidemiology*” (2010).

Table II
Levels of evidence for the quality of the measurement property

Level of evidence	Rating	Criteria
Strong	+++ OR −−−	Consistent findings in multiple studies of good methodological quality OR in one study of excellent methodological quality
Moderate	++ OR −−	Consistent findings in multiple studies of fair methodological quality OR in one study of good methodological quality
Limited	+ OR −	One study of fair methodological quality
Conflicting	±	Conflicting findings
Indeterminate	?	Only studies of poor methodological quality

+ = positive rating, ? = unknown rating, − = negative rating.
Adapted from Terwee *et al.* J Clin Epidemiol 2007;60(1):34–42.

the original, including Korean[16], Dutch[17,20–22], Austrian-German, Finnish, Norwegian, Portuguese, Spanish, Swedish[20] and Chinese[23].

Study characteristics such as cohort descriptors, sample sizes and instrument scores are provided in Table IV. The OA sites captured within the studies included hand, hip and knee[17,20], hip and knee[18], knee[23]or were not specified[16,19,21,22]. Stage or duration of OA was generally unreported. Participants were predominantly female across all studies and representative of the age of the wider OA population, with mean age ranging from 62 to 72.2 years.

Measurement property results and "best evidence synthesis"

Findings for measurement properties are summarised in Tables V and VI, qualitative data synthesis in Table VII.

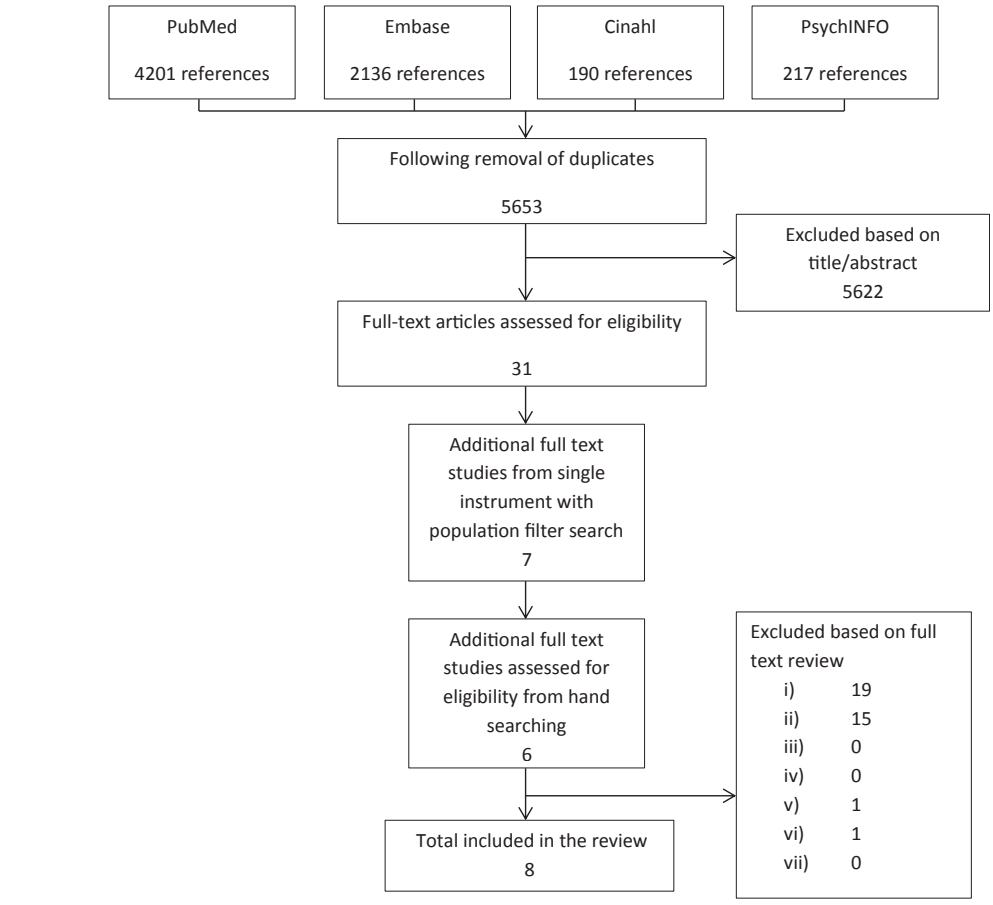


Fig. 1. Flowchart of the selection & inclusion of studies.

Table III
Content comparison of instruments measuring self-rated attitudes towards and capabilities to self-manage OA

Construct	Attitudes/beliefs pertaining to self-management of OA	Attitudes/beliefs pertaining to changing health behaviour	Knowledge required for self-management	Capability to perform skills required for self-management	Educational needs for self-management of OA	Interactions with health care providers assisting with management of OA	Overall capability to self-manage OA
MHLC[18]	✓						
PBC[19]	✓			✓		✓	
PAM-13[16]	✓		✓	✓		✓	✓
ENAT[20]	✓	✓	✓		✓		
PEPPI-5[21]			✓	✓		✓	
SCQOA[17]	✓	✓					
EC-17[22]	✓		✓	✓		✓	
PEPPI-10[23]			✓	✓		✓	

MHLC = Multidimensional Health Locus of Control, IHLC = Internal Health Locus of control, PBC = Perceived behavioural control, PAM-13 = Patient Activation Measure, ENAT = Educational needs assessment, PEPPI-5 = Perceived Efficacy in Patient–Physician Interactions Scale, SCQOA = The Stages of Change Questionnaire in Osteoarthritis, EC-17 = Effective Consumer Scale.
The ✓ indicated the presence of the construct of the instrument listed.

Internal consistency

Internal consistency was estimated for all instruments. Strong evidence (excellent rating) for internal consistency (Cronbach's $\alpha = 0.92$) was found for the Perceived Efficacy in Patient–Physician Interactions 5 item scale (PEPPI-5)[21], satisfying requirements for unidimensionality (Appendix 2). Moderate evidence (good rating) of adequate internal consistency was demonstrated for the Perceived Efficacy in Patient–Physician Interactions 10 item scale (PEPPI-10)[23] (Cronbach's $\alpha = 0.91$). Limited evidence (fair rating) of adequate internal consistency was found for three instruments: Perceived Behavioural Control (PBC)[19], PAM-13[16] and The Stages of Change Questionnaire in Osteoarthritis (SCQOA)[17]. There was indeterminate evidence (poor rating) of internal consistency for three instruments: Multidimensional Health Locus of Control (MHLC) (form C)[18], Educational Needs Assessment Tool (ENAT)[20] and Effective Consumer Scale (EC-17)[22].

Reliability

Adequate test-retest reliability required intraclass correlation coefficient (ICC) >0.7 (see Appendix 2). There was limited evidence (fair rating) of inadequate test-retest reliability for the PEPPI-5 (ICC = 0.68)[21]. Indeterminate evidence (poor rating) of adequate test–retest reliability was found for the EC-17[22] (ICC = 0.71).

Measurement error

Although data for test–retest reliability can be used to calculate measurement error, only one study reported this. There was indeterminate evidence of measurement error for the PEPPI-5[21] (limits of agreement (LOA) −6.83 to 6.35) because the MIC was not defined (see Appendix 2).

Structural validity

To demonstrate adequate structural validity, the factors identified should explain at least 50% of the variability of responses (see Appendix 2). There was strong evidence (excellent rating) that the PEPPI-5 featured an appropriate 1-factor structure[21]. There was moderate evidence (good rating) that the PEPPI-10 demonstrated a two factor structure[23]. There was limited evidence (fair rating) of positive structural validity for the PAM[16] and limited evidence (fair rating) that the factor structure of the SCQOA did not explain 50% of the variance[17]. There was also limited evidence (fair rating) of a negative result for structural validity of the ENAT[20]. The level of evidence for the structural validity of the EC-17, MHLC and PBC[18,19,22] was indeterminate (poor rating).

Hypothesis testing

The demonstration of adequate construct validity through hypothesis testing required that specific hypotheses were formulated *a priori* AND at least 75% of the results were in accordance with these[15]. There was limited evidence (fair rating) for adequate construct validity for the PEPPI-5[21] which was evaluated against; General Self Efficacy scale (GSES), Arthritis Impact Measurement Scales 2 Family and Friends scale (AIMS2 F & F), Short Form 36 mental component summary score, and pain numerical rating score (NRS). The EC-17 was compared with the same instruments as the PEPPI-5, however there was indeterminate evidence (poor rating) for the hypotheses tested (see Table IV)[22]. The study assessing PEPPI-10 did not formulate *a priori* hypotheses therefore the evidence for hypotheses testing was indeterminate[23].

Cross-cultural validity

Cross-cultural validity is established following specified translation procedures, then comparison of two cohorts differing only in language/cultural background to test if the translated instrument accurately reflects the measurements made in the original[12]. There was limited evidence (fair rating) for adequate translation of the English PAM[24] into Korean[16]. The Korean PAM was not compared with the English version. There was indeterminate evidence (poor rating) for the translation of the English EC-17[25] into Dutch[22] and no formal cross-cultural validation. There was limited evidence (fair rating) of adequate translation of the English PEPPI-10[26] into Chinese[23] with no cross-cultural validation. Cross-cultural comparisons were not made for the ENAT because the structural validity was inadequate in the OA group[20].

Floor and ceiling effects

Floor and ceiling effect results were rated using the quality criteria for rating the results of measurement properties in Appendix 2. There was strong evidence of absence of floor and ceiling effects for the PEPPI-5[21], limited evidence of a ceiling effect for the PEPPI-10[23] and indeterminate evidence for floor and ceiling effect for the EC-17[22].

Best evidence synthesis

The instrument with the most promising level of evidence for the measurement properties available was the PEPPI-5. Of note is that these results are applicable only to the Dutch language version of the PEPPI-5. There was strong evidence for internal consistency, structural validity, and lack of floor/ceiling effects, however there

Table IV
Characteristics of included studies of instruments measuring attitudes toward and/or capabilities regarding self-management of OA

Authors/ Instrument	Construct described	Time to administer	Availability	Language & country	Number, type of questions & scoring	Proportion with OA (%)	OA site & stage	% Other diseases in sample	N with >80% OA (response rate %)	Age: mean age years (SD) or age groups (%)	Female %	Mean (standard deviation), possible score range, distribution
Kelly (2007)/ MHLC[18]	Measures beliefs about who or what controls the patient's health status	Not stated	Freely available at: http://www.nursing.vanderbilt.edu/faculty/kwallston/mhlcscales.htm	English, USA & Canada	Three scales of six items each, using 6-point Likert scale measuring the following dimensions: "Internal" "Chance" and "Powerful Others". Sum the individual item scores for each subscale.	86.2	Hip & knee	Control sample: 13.8	1040 (100)	Study I: 65 (9) Study II: 64 (16) Study III: 62 (6)	Study I: (66) Study II: (59) Study III: (63)	IHLC: 26.44 (5.61) PHLC: 20.22 (6.64) CHLC: 16.96 (6.05) Each subscale has range 6–36
Liu (2007)/PBC[19]	Survey of OA patients' drug information seeking from physicians and pharmacists.	Not stated	In published paper	English USA	Eight statements with 7-point Likert responses Perceived difficulty: 3 Self-efficacy: 3 Controllability: 2 Answer for physicians & pharmacists separately	100	Not stated	–	1000 (61.9)	18–24: 1.8% 25–34: 3.8% 35–44: 11.9% 45–54: 27.6% 55–64: 28.3% >64: 26.6%	72.8	PDP: 5.10 (1.60) PDPh: 5.27 (1.49) SEP: 5.62 (1.62) SEPh: 5.62 (1.60) CP: 5.63 (1.36) CPh: 5.62 (1.37)
Ahn (2015)/PAM-13[16]	Patient activation: patient's knowledge, skill, and confidence regarding the self-management of a chronic disease	Not stated	Insignia health provides licenses for the PAM at a cost	Korean, South Korea	13-statements, with responses on a 4-point Likert scale. Raw score: sum responses to the 13 items. Scores ranging from 13 to 52. converted to a 0–100 interval scale. Higher total PAM scores reflect higher levels of patient activation.	100	Not stated	–	270 (100)	72.2 (8.3)	82.4	50.0 (13.5) 0–100
Ndosi (2014)/ ENAT[20]	Assesses the educational needs (priorities) of patients with rheumatic diseases	Not stated	Contact authors	Austrian German Finnish Dutch Norweigian Portuguese Spanish Swedish Austria Finland Netherlands Norway Portugal Spain Sweden	39 items with 4-point Likert scale in seven domains: managing pain (six items), movement (five items), feelings (four items), arthritis process (seven items), treatments (seven items), self-help measures (six items) and support systems (four items)	14.4	Hand, hip or knee in discussion. Stage not stated	AS: 22.5% FM: 12% PsA: 26.8% SLE: 12.3% SS: 12.0%	433 (response rate not stated)	Not stated for OA sample: pooled sample is 52.6 (13.1)	Not stated for OA but across pooled sample 1111111111112066.2	Not stated for OA group
ten Klooster (2012)/PEPPI-5[21]	Self-efficacy in both obtaining medical information and attention to chief health concern from a physician	Not stated	Dutch version freely available on web. English version published	Dutch, Nether-lands	Five questions with responses on a 5-point numerical rating scale. Total scores are summed to range from 5 to 25, higher total scores reflect higher perceived self-efficacy in patient –physician interactions.	100	Not stated	–	224 (55.4)	62.9 (10.2)	81.3	18.8 (4.3) 5–25 Slightly negatively skewed
Heuts (2005)/ SCQOA[17]	People move from low to high level of participation. Stages: no intention to change to optimal active participation with internalization of new behavior	3–5 min	Published in paper as appendix (in English)	Unclear (Dutch or English), Netherlands	21 items scored on 5-point Likert scale. Three subscales: seven questions for precontemplation, seven for contemplation, seven for action.	100	In results hip, knee & hand. Stage not stated	–	273 (100)	Range 40–60 years for inclusion criteria	59.7	Using highest score method: 10.3% was in the 'pre-contemplation stage', 22.3% in the 'contemplation stage', 67.0% was 'in action'
ten Klooster (2013)/EC-17[22]	Measures knowledge, attitudes, and behaviours regarding self-management skills	Not stated	Available in published paper & on web http://www.cgh.uottawa.ca/assets/documents/Survey.pdf	Dutch Netherlands	17 items with 5-point Likert scale. Item scores are summed when items are completed and converted to range from 0 to 100, where 100 is the best possible score.	85.6	Not stated	FM: 14.4	209 (55.8% of combined OA & FM sample)	62.6 (10.1)	80.9	68.9 (16.3), 0–100, near normal distribution (Kolmogorov –Smirnov, $P = 0.058$)

(*continued on next page*)

Table IV (*continued*)

Authors/ Instrument	Construct described	Time to administer	Availability	Language & country	Number, type of questions & scoring	Proportion with OA (%)	OA site & stage	% Other diseases in sample	N with >80% OA (response rate %)	Age: mean age years (SD) or age groups (%)	Female %	Mean (standard deviation), possible score range, distribution
Zhao (2016) PEPPI-10[23]	Self-efficacy in both obtaining medical information and attention to chief health concern from a physician	Not stated	Supplement link from paper: https://www.dovepress.com/get_supplementary_file.php?f=110883.pdf	Chinese, China	10 items with 10 point numerical rating scale: Not confident to extremely confident. Sum ten scores from 0 to 100 (100 best self-efficacy)	100	Knee	–	115 (100)	63.42 (6.7)	59	90.07 (12.9), 0 –100 Negatively skewed distribution

IHLC = Internal health locus of control, PHLC = powerful others health locus of Control, CHLC = chance health locus of control, PDP = perceived difficulty for physicians, PDPh = perceived difficulty for pharmacists, SEP = self-efficacy for physicians, SEPh = self-efficacy for pharmacists, CP = controllability for physicians, CPh = controllability for pharmacists, PAM-13 = Patient Activation Measure-13, RA: rheumatoid arthritis, FM: fibromyalgia, AS: ankylosing spondylitis, PsA: psoriatic arthritis, SLE: systemic lupus erythematosus, SS: systemic sclerosis.

was limited positive evidence for construct validity (hypothesis testing) and limited evidence of negative findings for test-retest reliability (Tables VI and VII). There was indeterminate evidence for measurement error and no information for content validity, or responsiveness.

Discussion

OA self-management programmes are not curative, but aim to equip participants with the tools to manage their disease. It is important to measure the changes in attitudes towards and/or capabilities regarding OA self-management to determine whether participants achieve this aim and to demonstrate efficacy of programmes. Further, it may be possible to predict outcomes of participants by measuring attitudes towards and/or capabilities in regards to OA self-management at baseline. This may provide a basis on which to appropriately allocate healthcare resources to those that will likely benefit from such a programme. Participants reporting a positive attitude toward self-management and good self-management capabilities may be prioritised for immediate engagement in a programme. Conversely, individuals reporting poorer attitudes and capabilities may be targeted for supplementary therapies such as motivational coaching to improve the likelihood of successful participation in such a programme. In order to test whether this is possible, we first need to identify a suitable instrument measuring attitudes towards and/or abilities regarding self-management of OA that demonstrates good measurement properties.

This systematic review is the first to synthesize the measurement property evidence for instruments assessing attitudes towards and/or capabilities regarding self-management of OA. There were a very small number of studies identified; only eight studies reported measurement properties of such instruments, each for a separate instrument. The scope of measurement properties assessed by the included studies was very limited. Internal consistency and structural validity was estimated for all instruments. Test-retest reliability[21,22], and hypothesis testing[21,22] were each assessed for two instruments, cross-cultural validity was addressed in three studies[16,22,23]. Measurement error was reported in one study[21], responsiveness and content validity were not evaluated for any of the instruments.

Given the limited measurement property evidence for the included instruments we cannot provide a definitive, evidence-based recommendation for a particular instrument to measure attitudes towards and capabilities regarding OA self-management on the basis of good measurement properties. On balance, the instrument with the "best" measurement properties was the Dutch version of the PEPPI-5[21]. There was strong evidence that the PEPPI-5 satisfied requirements for internal consistency and structural validity. There was limited evidence for the hypotheses specified comparing PEPPI-5 scores against several other PROMs. The test–retest reliability findings were sub-optimal (i.e., ICC < 0.7) which has implications regarding the standard error of the measure. Greater standard error may require larger change scores to represent 'real' change (vs error inherent in the measure) between groups over time. The evidence for measurement error of the PEPPI-5 was indeterminate because the MIC was not provided. Measurement property evidence for content validity and responsiveness of the PEPPI-5 remains unknown. The remaining instruments identified in the review demonstrated moderate evidence of positive measurement properties at best.

The PEPPI-5 was originally developed in a sample of "older people" with mixed medical diagnoses; measurement property results for internal consistency, structural and construct validity were reported for this population[26]. Given the PEPPI-5 was developed for a different group of patients it may be that it has limited

Table V
Measurement properties instruments measuring self-management of OA according to the COSMIN checklist with 4-point scale: internal consistency, reliability, measurement error and structural validity

Instrument	*Requirements IRT	Internal consistency		Reliability		Measurement error		Structural validity	
		Result Cronbach's alpha	COSMIN score	Result	COSMIN score	Result	COSMIN score	Result	COSMIN score
MHLC[18]	Good	IHLC: 0.75; PHLC: 0.70; CHLC: 0.65	Poor	–	–	–	–	Confirmatory FA, three factor model: $\chi^2 = 904.50$, 135 df, ($P < 0.01$), RMSEA 0.0, GFI = 0.96, CFI = 0.79, ECVI = 0.81, PCA, FA & Rasch analysis supported item reduction: removed two items	Poor
PBC[19]	–	PDP: $\alpha = 0.77$ PDPh: $\alpha = 0.72$ SEP:$\alpha = 0.83$ SEPh: $\alpha = 0.83$	Fair	–	–	–	–	PCA & exploratory FA with Factor loading. Data reduction & data detection	Fair
PAM-13[16]	Good	$\alpha = 0.88$	Fair	–	–	–	–	Confirmatory PCA GFI = 32 (11.9%) misfits MNSQ 0.68 to 1.42 Rasch analysis: person reliability was between .87 (real) and .89 (model), and the item reliability was 0.99. The separation index for persons was 2.57 and that for items was 10.56 57.5% variance of data explained	Fair
ENAT[20]	Good	IRT: Person separation index >0.9	Poor	–	–	–	–	Confirmatory FA, structure detection & Rasch analysis OA group was a misfit	Fair
PEPPI-5[21]	–	$\alpha = 0.92$	Excellent	Test-retest: ICC 0.68 (95% CI 0.56, 0.78) Bland–Altman analysis LOA 6.83–6.35 (mean difference −0.24, t(99) = −0.71, $P = 0.48$)	Fair	LOA −6.83 to 6.35 differences − weakly related to the magnitude of the measurement ($r2 = 0.04$, $P = 0.049$), indicating little to no systematic bias	Fair	Confirmatory FA, factor loading & structure detection (1 factor) SB χ^2 (5) = 17.43, NNFI = 0.98, CFI = 0.99, SRMR = 0.03, RMSEA (90% CI) = 0.11 (0.05 −0.16)	Excellent
SCQOA[17]	–	Action $\alpha = 0.74$ Precontemplation $\alpha = 0.70$ Contemplator $\alpha = 0.77$ After removal of 5 items: Action $\alpha = 0.79$ Precontemplation $\alpha = 0.72$ Contemplation $\alpha = 0.76$	Fair	–	–	–	–	Confirmatory FA, factor loading & date reduction: removal of items 3, 7, 12, 16, 18 and 20 PCA Repeated FA with 15 item scale: 3 factors explained 45% of variance	Fair
EC-17[22]	Good	Person reliability: 0.92	Poor	Test-retest ICC = 0.71 (95% CI: 0.60 −0.80)	Poor	–	–	Confirmatory FA Apart from RMSEA, 1-factor model good fit SB χ^2 (119) = 488.70, NNFI = 0.96, CFI = 0.96, SRMR = 0.08, RMSEA (90% CI) = 0.11 (0.10 −0.12)	Poor
PEPPI-10[23]	–	$\alpha = 0.91$	Good	–	–	–	–	Confirmatory FA: two-factor model good fit (df = 33, P-value = 0.000) except RMSEA = 0.164 above cutoff	Good

Note: Content validity, criterion validity and responsiveness were not reported on in any included articles, hence do not appear in the table.
IHLC = internal health locus of control, PHLC = powerful others health locus of control, CHLC = chance health locus of control, PDP = perceived difficulty for physicians, PDPh = perceived difficulty for pharmacists, SEP = self-efficacy for physicians, SEPh = self-efficacy for pharmacists, CP = controllability for physicians, CPh = controllability for pharmacists, PAM-13 = patient Activation Measure-13. FA = factor analysis, PCA = principal components analysis, GFI = goodness of fit index, MNSQ = infit & outfit mean square statistics, NRS = numerical rating score, NS = non-significant, NNFI = non-normed fit index, CFI = comparative fit index, SRMR = standardized root mean square residual, RMSEA = root mean square error of approximation, SB χ^2 = Satorra–Bentler chi-squared statistic, LOA = limits of agreement, MFES = modified fall efficacy scale, OSES = osteoporosis self-efficacy scale, SEE-C = self-efficacy for exercise scale.
* This field was only completed for those instruments based on Item Response Theory (IRT).

Table VI
Measurement properties of instruments measuring self-management: construct validity, cross-cultural validity, and floor and ceiling effects

Instrument	Construct validity (Hypothesis testing)			Cross-cultural validity		Floor & ceiling effects
	Hypothesis	Result	COSMIN score	Result	COSMIN score	Result
MHLC[18]	–	–	–	–	–	Seven items, including all six items of the IHLC scale, exhibited skewness that exceeded −1.00 (i.e., a "ceiling effect"). No floor effect
PBC[19]	–	–	–	–	–	–
PAM-13[16]	–	–	–	Items 1 and 4 were adjusted to make more sense in Korean translation. PCA indicated unidimensionality	Fair*	–
ENAT[20]	–	–	–	–	–	–
PEPPI-5[21]	Expected correlations: Strongly positively correlated with EC-17, moderately positively with GSES, weakly positively with AIMS2 family & friends scale and SF-36 MCS and not correlated with SF-36 PCS and pain NRS	EC-17: $r = 0.52$, $P < 0.01$ GSES: $r = 0.07$ (not sig) AIMS2 F & F: $r = 0.23$, $P < 0.05$ SF-36 MCS: $R = 0.26$, $P < 0.01$ SF-36 PCS: $r = 0.05$ (NS) Pain NRS: $r = -0.12$ (NS)	Fair	–	–	No floor and ceiling effects: no patients scored five and 26 patients (11.6%) scored 25
SCQOA[17]	–	–	–	–	–	–
EC-17[22]	Expected correlations: Strongly correlated PEPPI-5, moderately correlated GSES and AIMS2 F & f, moderate correlation SF-36 MCS, weak correlations SF-36 PCS & pain NRS	PEPPI-5: $r = 0.55$, $P < 0.01$ GSES: $r = 0.26$, $P < 0.01$ AIMS2 F & F: $r = -0.34$, $P < 0.01$ SF-36 MCS: $r = 0.39$, $P < 0.01$ SF-36 PCS: $r = 0.14$, $P < 0.05$ Pain NRS: $r = -0.21$, $P < 0.01$	Poor	Following pretests small wording changes made in six items. CFA supported unidimensional structure of the scale	Poor*	No ceiling or floor effect found: no participants scored zero and only 1.3% achieved maximum score
PEPPI-10[23]	No hypothesis and expected correlations not stated	SEE-C: $r = 0.292$, $P < 0.01$ MFES: $r = 0.220$, $P < 0.05$ OSES: $r = 0.315$, $P < 0.01$	Poor	Following pretests, two items were modified to suit Chinese language. FA showed Chinese version of PEPPI-10 has two common factors; different to 1 factor reported previously for the English version	Fair*	Ceiling effect found for 28.2% of participants. No floor effect

Note: Content validity, criterion validity and responsiveness were not reported on in any included articles, hence do not appear in the table. Floor and ceiling effects were not evaluated using the COSMIN Checklist.
IHLC = internal health locus of control, PAM-13 = Patient Activation Measure-13, GSES = General Self Efficacy scale, AIMS2 F & F = Dutch Arthritis Impact Measurement Scales 2 Family and Friends scale, SF-36 MCS = short form 36 mental component summary score, SF-36 PCS = short form 36 mental component summary score, MFES = modified fall efficacy scale, OSES = osteoporosis self-efficacy scale, SEE-C = self-efficacy for exercise scale.
* Paper did not assess cross-cultural validity however did translate the questionnaire into other language(s) hence quality of translation items of COSMIN checklist were rated (Box G items 4–11).

content validity for OA. The PEPPI-5 measures self-efficacy in obtaining both medical information and attention to chief health concern from a physician, hence includes limited aspects of a patient's ability to self-manage OA. Although effective communication with a physician is important, it may not be a key outcome used to indicate the efficacy of such programmes. OA self-management programmes are often multidisciplinary, with input from a team of health professionals including physiotherapists, dietitians and occupational therapists[27], and some programmes do not include a medical physician[28]. Hence, there is a clear need to develop tools that have adequate content validity for participants of OA self-management programmes.

A previous systematic review synthesized the measurement property evidence for instruments measuring self-efficacy in participants with rheumatic conditions[29]. Self-efficacy is defined as the confidence that one possesses the ability to influence events that affect aspects of one's life[30]. Self-efficacy is potentially an important aspect of self-management, however additional constructs may be considered such as how motivated or activated participants are to self-manage[24], or beliefs about who controls their health[18].

Table VII
Summary of the assessment of measurement properties of all instruments using COSMIN rating, quality criteria for rating the results of measurement properties and levels of evidence

Instrument	Internal consistency	Reliability	Measurement error	Structural validity	Hypothesis testing	Cross-cultural validity	Floor and ceiling effects
MHLC[18]	?	0	0	?	0	0	?
PBC[19]	+	0	0	?	0	0	0
PAM-13[16]	+	0	0	+	0	*+	0
ENAT[20]	?	0	0	–	0	0	0
PEPPI-5[21]	+++	–	?	+++	+	0	+++
SCQOA[17]	+	0	0	–	0	0	0
EC-17[22]	?	?	0	?	?	*?	?
PEPPI-10[23]	++	0	0	++	?	*+	–

Note: Content validity and responsiveness were not reported on in any included studies, hence do not appear in the table.
+++ or −−− strong evidence, ++ or −− moderate evidence, + or − limited evidence, ± conflicting evidence, ? indeterminate, 0 no information [+ positive, − negative rating (results)]. IHLC = internal health locus of control, PAM-13 = Patient Activation Measure-13.
* Paper did not assess cross-cultural validity hence the quality criteria for rating the results of measurement properties (Appendix 2) were not applied to the overall measurement property result, however the translation items of COSMIN checklist were rated (Box G items 4–11).

The previous review included participants of mixed disease groups with different rheumatic conditions[29]. Given that measurement property evidence is specific to the population studied, these measurement property results cannot be extrapolated to the OA population. The population-specific nature of measurement properties also placed limitations on the studies available for this current review. Often studies were excluded at the full-text stage because they comprised mixed disease cohorts and did not report the OA participant results separately. This limited the number of studies included.

The methodologies of the included studies were limited to investigation of a small range of measurement properties. Internal consistency and structural validity were reported for all studies. This is similar finding to the previous systematic review of self-efficacy in patients with rheumatic conditions[29]. Although these are valuable measurement properties to establish, many measurement properties remain untested in the instruments of our systematic review. Test–retest reliability estimates the relative consistency of a measure in otherwise stable patients, so that when any change is detected by the instrument, it can be attributed to the intervention rather than from measurement error of the instrument. Unfortunately the test–retest reliability and measurement error for the included instruments are yet to be established in OA patients. Test–retest reliability was tested in a larger proportion of studies included in the systematic review on rheumatic conditions, however the quality of the evidence was generally poor and measurement error was unreported[29]. Hypothesis testing is a further property that was neglected by the majority of studies in our review. Hypothesis testing establishes whether an instrument measures the intended construct by testing the internal relationships with scores of other instruments measuring similar or different constructs[13]. There is much need for future studies evaluating test–retest reliability, measurement error and construct validity of instruments measuring OA self-management attitudes and capabilities.

Cross-cultural validation was attempted in three studies that translated questionnaires; however, true cross-cultural validation comparing language versions was not conducted. This was also found in the previous review of instruments measuring self-efficacy[29]. We found no evidence pertaining to content validity, responsiveness, or MID/MIC. Similar to previous conclusions[29], the recommendations arising from the present review are limited due to the small number of studies, their poor methodology, and the limited scope of measurement properties assessed. Further studies concerned with all measurement properties of existing instruments assessing self-management of OA is the only way to remedy this situation.

Some existing instruments measuring attitudes towards and/or capabilities regarding OA self-management were not featured in the systematic review because there was no measurement property evidence available. The Health Education Impact Questionnaire (heiQ)[31] evaluates the efficacy of patient education programs and has been used to evaluate OA self-management programs[5,32]. Also, the Arthritis Self Efficacy Score (ASES) measures patients' perceived self-efficacy to cope with the symptoms and limitations attributed to chronic arthritis[33] and is a published outcome of existing OA self-management programs[34,35]. The measurement properties of the heiQ and ASES remain untested in the OA population. Given the current popularity of these instruments, the measurement properties of heiQ and ASES are an important area of future research.

There were possible limitations of this systematic review; the inclusion criteria requiring studies to be published as original articles may have introduced publication bias. Unpublished studies may have been more likely to contain evidence of negative results about measurement properties of the instruments under study. However, the inclusion of only peer-reviewed articles likely enhanced the quality of included studies, given the basic level of scrutiny required to publish. This may have improved the quality of the review rather than biasing it. While excluding non-English language studies may have introduced bias, no such studies were identified by the comprehensive search strategy.

Conclusion

This review highlights the paucity of evidence available for the measurement properties of instruments assessing attitudes towards and/or capabilities regarding OA self-management. There were many gaps in the measurement property evidence for the instruments identified. The instrument with the “best” properties assessed self-efficacy in communication with a physician; a very discrete aspect of self-management. Therefore, we were unable to make recommendations concerning instruments to assess attitudes toward and/or capabilities regarding OA self-management. Further well-designed studies of measurement properties of available instruments are required. This review may provide a starting point for researchers to identify the instruments that are currently used for this purpose in the OA population and the evidence for measurement properties available. Once we are able to identify instruments with adequate measurement properties for use in this population, we will be able to better compare the efficacy of different OA self-management programmes and inform best practice for care of our patients.

Author contributions

JPE, KM, and DH conceived the study, JPE, KM, DH, FD, NC, SM and BRL contributed to the study design. JPE and KM developed the search strategy and performed the literature search, JPE, SM and KM screened the abstracts for eligibility. JPE, SM, NC, KM and FD performed and contributed to the quality ratings. JPE, SM and KM extracted data. JE wrote the manuscript, KM, DH, FD, NC and BRL edited the manuscript. All authors read and approved the final manuscript.

Conflict of interest

There are no conflicts of interest to declare.

Role of the funding source

There was no funding source for this study.

Acknowledgements

The authors wish to thank Jeremy Cullis, Clinical Librarian Macquarie University, Sydney, Australia. Jeremy generously contributed his expertise as a research librarian to assist in building the comprehensive search strategy and assisted greatly in translating the measurement property filter into the language of the different databases.

Appendix 1. Search strategy

i) Construct

generalized self efficacy scale[tiab] OR adaptive behavior[tiab] OR multidimensional health locus of control[tiab] OR pain self efficacy questionnaire[tiab] OR health literacy management scale [tiab] OR stages of change questionnaire in osteoarthritis[tiab] OR health education impact questionnaire[tiab] OR patient activation measure[tiab] OR effective consumer scale[tiab] OR arthritis self-efficacy scale[tiab] OR internal-external control[MH] OR locus of control[tw] OR attitude to health[MH] OR health locus of control

[tiab] OR adaptation, psychological[MH] OR health behavior[MH] OR health knowledge, attitudes, practice[MH] OR self management behavio*[tiab] OR patient activation[tiab] OR self concept [MH] OR self efficacy[MH] OR confidence[tiab] OR activation[tiab] OR consumer participation[MH] OR patient education as topic [MH] OR Patient Participation[MH] OR individualized medicine [MH] OR patient-centered care[MH] OR goals[MH] OR patient preference[MH] OR choice behavior[MH] OR decision making [MH] OR patient care planning[MH] OR personalised care planning[tiab] OR patient led[tiab] OR selftreatment[tiab] OR self treat*[tiab] OR self manage*[tiab] OR self care[tiab] OR self care [MH]

ii) Target population

osteoarthritis[MH] OR osteoarth*[tiab] OR degenerative arthritis [tiab] OR arthrosis[tiab]

iii) Measurement instrument filter

instrumentation[sh] OR methods[sh] OR validation studies[pt] OR Comparative Study[pt] OR psychometrics[MH] OR psychometr* [tiab] OR clinimetr*[tw] OR clinometr*[tw] OR "outcome assessment (health care)"[MH] OR "outcome assessment"[tiab] OR "outcome measure*"[tw] OR "observer variation"[MH] OR "observer variation"[tiab] OR "Health Status Indicators"[MH] OR "reproducibility of results"[MH] OR reproducib*[tiab] OR "discriminant analysis"[MH] OR reliab*[tiab] OR unreliab*[tiab] OR valid*[tiab] OR coefficient[tiab] OR homogeneity[tiab] OR homogeneous[tiab] OR "internal consistency"[tiab] OR (cronbach*[tiab] AND (alpha[tiab] OR alphas[tiab])) OR (item[tiab] AND (correlation*[tiab] OR selection*[tiab] OR reduction*[tiab])) OR agreement [tiab] OR precision[tiab] OR imprecision[tiab] OR "precise values"[tiab] OR test–retest[tiab] OR (test[tiab] AND retest[tiab]) OR (reliab*[tiab] AND (test[tiab] OR retest[tiab])) OR stability[tiab] OR interrater[tiab] OR inter-rater[tiab] OR intrarater[tiab] OR intrarater[tiab] OR intertester[tiab] OR inter-tester[tiab] OR intratester [tiab] OR intra-tester[tiab] OR interobserver[tiab] OR inter-observer [tiab] OR intraobserver[tiab] OR intraobserver[tiab] OR intertechnician[tiab] OR inter-technician[tiab] OR intratechnician[tiab] OR intra-technician[tiab] OR interexaminer[tiab] OR interexaminer[tiab] OR intraexaminer[tiab] OR intra-examiner[tiab] OR interassay[tiab] OR inter-assay[tiab] OR intraassay[tiab] OR intra-assay[tiab] OR interindividual[tiab] OR inter-individual[tiab] OR intraindividual[tiab] OR intra-individual[tiab] OR interparticipant[tiab] OR inter-participant[tiab] OR intraparticipant [tiab] OR intra-participant[tiab] OR kappa[tiab] OR kappa's[tiab] OR kappas[tiab] OR repeatab*[tiab] OR ((replicab*[tiab] OR repeated [tiab]) AND (measure[tiab] OR measures[tiab] OR findings[tiab] OR result[tiab] OR results[tiab] OR test[tiab] OR tests[tiab])) OR generaliza*[tiab] OR generalisa*[tiab] OR concordance[tiab] OR (intraclass[tiab] AND correlation*[tiab]) OR discriminative[tiab] OR "known group"[tiab] OR factor analysis[tiab] OR factor analyses [tiab] OR dimension*[tiab] OR subscale*[tiab] OR (multitrait[tiab] AND scaling[tiab] AND (analysis[tiab] OR analyses[tiab])) OR item discriminant[tiab] OR interscale correlation*[tiab] OR error[tiab] OR errors[tiab] OR "individual variability"[tiab] OR (variability[tiab] AND (analysis[tiab] OR values[tiab])) OR (uncertainty[tiab] AND (measurement[tiab] OR measuring[tiab])) OR "standard error of measurement"[tiab] OR sensitiv*[tiab] OR responsive*[tiab] OR ((minimal[tiab] OR minimally[tiab] OR clinical[tiab] OR clinically [tiab]) AND (important[tiab] OR significant[tiab] OR detectable [tiab]) AND (change[tiab] OR difference[tiab])) OR (small*[tiab] AND (real[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR meaningful change[tiab] OR "ceiling effect"[tiab] OR "floor effect"[tiab] OR "Item response model"[tiab] OR IRT[tiab] OR Rasch[tiab] OR "Differential item functioning"[tiab] OR DIF[tiab] OR "computer adaptive testing"[tiab] OR "item bank"[tiab] OR "cross-cultural equivalence"[tiab]

iv) Exclusion filter

"addresses"[PT] OR "biography"[PT] OR "case reports"[PT] OR "comment"[PT] OR "directory"[PT] OR "editorial"[PT] OR "festschrift"[PT] OR "interview"[PT] OR "lectures"[PT] OR "legal cases"[PT] OR "legislation"[PT] OR "letter"[PT] OR "news"[PT] OR "newspaper article"[PT] OR "patient education handout"[PT] OR "popular works"[PT] OR "congresses"[PT] OR "consensus development conference"[PT] OR "consensus development conference, nih"[PT] OR "practice guideline"[Publication Type]) NOT ("animals"[MeSH Terms] NOT "humans"[MeSH Terms]

Appendix 2. Quality criteria for rating the results of measurement properties

Measurement property	Rating	Quality criteria
Internal Consistency	+	Factor analyses performed on adequate sample size (7*# items and >100) AND Cronbach's alpha(s) calculated per dimension AND Cronbach's alpha(s) between 0.70 and 0.95
	?	No factor analysis OR doubtful design or method
	−	Cronbach's alpha(s) ≤0.70 or ≥0.95, despite adequate design and method
	0	No information found on internal consistency
Reliability	+	ICC or weighted Kappa >0.70
	?	Doubtful design or method (e.g., time interval not mentioned)
	−	ICC or weighted Kappa <0.70, despite adequate design and method
	0	No information found on reliability
Measurement error	+	MIC > SDC OR MIC outside the LOA
	?	MIC not defined or doubtful design
	−	MIC < SDC OR MIC equals or inside LOA
	0	No information found on measurement error
Structural validity	+	Factors should explain at least 50% of the variance
	?	Explained variance not mentioned
	−	Factors explain <50% of the variance
	0	No information found on structural validity
Hypothesis testing	+	Specific hypotheses were formulated AND at least 75% of the results are in accordance with these hypotheses
	?	Doubtful design or method (e.g., no hypotheses)
	−	Less than 75% of hypotheses were confirmed, despite adequate design and methods
	0	No information found on hypothesis testing
Cross-cultural validity	+	Original factor structure confirmed or no important DIF found between language versions
	?	Confirmatory factor analysis not applied & DIF not assessed
	−	Original factor structure not confirmed or important DIF found between language versions
	0	No information found on cross-cultural validity
Floor and ceiling effects	+	≤15% of the respondents achieved the highest or lowest possible scores
	?	Doubtful design or method
	−	>15% of the respondents achieved the highest or lowest possible scores despite adequate design and methods
	0	No information found on interpretation

ICC = intraclass correlation coefficient, LOA = limits of agreement, MIC = minimal important change, SDC = smallest detectable change.
Adapted from Terwee *et al.* J Clin Epidemiol 2007; 60(1): 34–42 and F. Dobson *et al.* Osteoarthritis and Cartilage 20 (2012) 1548–1562. Content and criterion validity, responsiveness, & interpretability were not reported on in any included studies; hence have been omitted.

References

1. Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, *et al*. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2013;380(9859):2163–96. PubMed PMID: 23245607.
2. Epping-Jordan JE, Pruitt SD, Bengoa R, Wagner EH. Improving the quality of health care for chronic conditions. Qual Saf Health Care 2004;13(4):299–305, http://dx.doi.org/10.1136/qhc.13.4.299. Epub 2004/08/04. PubMed PMID: 15289634; PubMed Central PMCID: PMCPmc1743863.
3. Coulter A, Entwistle VA, Eccles A, Ryan S, Shepperd S, Perera R. Personalised care planning for adults with chronic or long-term health conditions. Cochrane Database Syst Rev 2015;(3):Cd010523, http://dx.doi.org/10.1002/14651858.CD010523.pub2. Epub 2015/03/04. PubMed PMID: 25733495.
4. Du S, Yuan C, Xiao X, Chu J, Qiu Y, Qian H. Self-management programs for chronic musculoskeletal pain conditions: a systematic review and meta-analysis. Patient Educ Couns 2011;85(3):e299–310. PubMed PMID: 21458196.
5. Kroon FP, van der Burg LR, Buchbinder R, Osborne RH, Johnston RV, Pitt V. Self-management education programmes for osteoarthritis. Cochrane Database Syst Rev 2014;1:Cd008963, http://dx.doi.org/10.1002/14651858.CD008963.pub2. Epub 2014/01/16. PubMed PMID: 24425500.
6. Tugwell P, Boers M, D'Agostino MA, Beaton D, Boonen A, Bingham 3rd CO, *et al*. Updating the OMERACT filter: implications of filter 2.0 to select outcome instruments through assessment of "truth": content, face, and construct validity. J Rheumatol 2014;41(5):1000–4, http://dx.doi.org/10.3899/jrheum.131310. Epub 2014/04/03. PubMed PMID: 24692531; PubMed Central PMCID: PMCPMC4212637.
7. Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for outcome measures in rheumatology. J Rheumatol 1998;25(2):198–9. Epub 1998/03/07. PubMed PMID: 9489805.
8. de Vet H, Terwee C, Mokkink L, Knol D. Measurement in Medicine: A Practical Guide to Biostatistics and Epidemiology. London: Cambridge University Press; 2011.
9. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. Patient Educ Couns 2002;48(2):177–87. Epub 2002/10/29. PubMed PMID: 12401421.
10. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009;151(4):264–9, http://dx.doi.org/10.7326/0003-4819-151-4-200908180-00135.
11. Terwee CB, Jansma EP, Riphagen II, de Vet HCW. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual Life Res 2009;18(8):1115–23, http://dx.doi.org/10.1007/s11136-009-9528-5. PubMed PMID: PMC2744791.
12. Mokkink L, Terwee C, Patrick D, Alonso J, Stratford P, Knol D, *et al*. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res 2010;19(4):539–49, http://dx.doi.org/10.1007/s11136-010-9606-8.
13. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, *et al*. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010;63(7):737–45, http://dx.doi.org/10.1016/j.jclinepi.2010.02.006. PubMed PMID: 20494804.
14. Terwee C, Mokkink L, Knol D, Ostelo RJG, Bouter L, de Vet HW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res 2012;21(4):651–7, http://dx.doi.org/10.1007/s11136-011-9960-1.
15. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, *et al*. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;60(1):34–42, http://dx.doi.org/10.1016/j.jclinepi.2006.03.012. Epub 2006/12/13. PubMed PMID: 17161752.
16. Ahn YH, Yi CH, Ham OK, Kim BJ. Psychometric properties of the Korean version of the "Patient Activation Measure 13" (PAM13-K) in patients with osteoarthritis. Eval Health Prof 2015;38(2):255–64, http://dx.doi.org/10.1177/0163278714540915. Epub 2014/07/06. PubMed PMID: 24986844.
17. Heuts PH, de Bie RA, Dijkstra A, Aretz K, Vlaeyen JW, Schouten HJ, *et al*. Assessment of readiness to change in patients with osteoarthritis. development and application of a new questionnaire. Clin Rehabil 2005;19(3):290–9. Epub 2005/04/30. PubMed PMID: 15859530.
18. Kelly PA, Kallen MA, Suarez-Almazor ME. A combined-method psychometric analysis recommended modification of the multidimensional health locus of control scales. J Clin Epidemiol 2007;60(5):440–7, http://dx.doi.org/10.1016/j.jclinepi.2006.08.005. Epub 2007/04/11. PubMed PMID: 17419954.
19. Liu Y, Doucette WR, Farris KB. Perceived difficulty and self-efficacy in the factor structure of perceived behavioral control to seek drug information from physicians and pharmacists. Res Soc Adm Pharm RSAP 2007;3(2):145–59, http://dx.doi.org/10.1016/j.sapharm.2006.07.002. Epub 2007/06/15. PubMed PMID: 17561217.
20. Ndosi M, Bremander A, Hamnes B, Horton M, Kukkurainen ML, Machado P, *et al*. Validation of the educational needs assessment tool as a generic instrument for rheumatic diseases in seven European countries. Ann Rheum Dis 2014;73(12):2122–9, http://dx.doi.org/10.1136/annrheumdis-2013-203461.
21. ten Klooster PM, Oostveen JC, Zandbelt LC, Taal E, Drossaert CH, Harmsen EJ, *et al*. Further validation of the 5-item Perceived Efficacy in Patient-Physician Interactions (PEPPI-5) scale in patients with osteoarthritis. Patient Educ Couns 2012;87(1):125–30, http://dx.doi.org/10.1016/j.pec.2011.07.017. Epub 2011/09/06. PubMed PMID: 21889864.
22. ten Klooster PM, Taal E, Siemons L, Oostveen JC, Harmsen EJ, Tugwell PS, *et al*. Translation and validation of the Dutch version of the Effective Consumer Scale (EC-17). Qual Life Res 2013;22(2):423–9, http://dx.doi.org/10.1007/s11136-012-0162-2. Epub 2012/03/29. PubMed PMID: 22453645; PubMed Central PMCID: PMCPmc3576564.
23. Zhao H, Luo W, Maly RC, Liu J, Lee J, Cui Y. Validation of the Chinese version 10-item perceived efficacy in patient-physician interactions scale in patients with osteoarthritis. Patient Prefer Adherence 2016;10:2189–95. PubMed PMID: 613179091.
24. Hibbard JH, Stockard J, Mahoney ER, Tusler M. Development of the Patient Activation Measure (PAM): conceptualizing and measuring activation in patients and consumers. Health Serv Res 2004;39(4 Pt 1):1005–26, http://dx.doi.org/10.1111/j.1475-6773.2004.00269.x. PubMed PMID: PMC1361049.
25. Kristjansson E, Tugwell PS, Wilson AJ, Brooks PM, Driedger SM, Gallois C, *et al*. Development of the effective musculoskeletal consumer scale. J Rheumatol 2007;34(6):1392–400.
26. Maly RC, Frank JC, Marshall GN, DiMatteo MR, Reuben DB. Perceived efficacy in patient-physician interactions (PEPPI): validation of an instrument in older persons. J Am Geriatrics Soc 1998;46(7):889–94. Epub 1998/07/22. PubMed PMID: 9670878.

27. Eyles JP, Lucas BR, Patterson JA, Williams MJ, Weeks K, Fransen M, *et al*. Does clinical presentation predict response to a nonsurgical chronic disease management program for end-stage hip and knee osteoarthritis? J Rheumatol 2014;41(11): 2223–31, http://dx.doi.org/10.3899/jrheum.131475.
28. Skou ST, Odgaard A, Rasmussen JO, Roos EM. Group education and exercise is feasible in knee and hip osteoarthritis. Dan Med J 2012;59(12):A4554. Epub 2013/01/08. PubMed PMID: 23290290.
29. Garratt AM, Lochting I, Smedslund G, Hagen KB. Measurement properties of instruments assessing self-efficacy in patients with rheumatic diseases. Rheumatology (Oxford, England) 2014;53(7):1161–71, http://dx.doi.org/10.1093/rheumatology/ket374. Epub 2013/11/20. PubMed PMID: 24249031.
30. Bandura A. Self-efficacy. The Corsini Encyclopedia of Psychology. John Wiley & Sons, Inc.; 2010.
31. Osborne RH, Elsworth GR, Whitfield K. The Health Education Impact Questionnaire (heiQ): an outcomes and evaluation measure for patient education and self-management interventions for people with chronic conditions. Patient Educ Couns 2007;66(2): 192–201, http://dx.doi.org/10.1016/j.pec.2006.12.002.
32. Umapathy H, Bennell K, Dickson C, Dobson F, Fransen M, Jones G, *et al*. The web-based osteoarthritis management resource my joint pain improves quality of care: a quasi-experimental study. J Med Internet Res 2015;17(7):e167, http://dx.doi.org/10.2196/jmir.4376. Epub 2015/07/15. PubMed PMID: 26154022; PubMed Central PMCID: PMCPmc4526979.
33. Lorig K, Chastain RL, Ung E, Shoor S, Holman HR. Development and evaluation of a scale to measure perceived self-efficacy in people with arthritis. Arthritis Rheumatism 1989;32(1): 37–44. Epub 1989/01/01. PubMed PMID: 2912463.
34. Skou ST, Simonsen ME, Odgaard A, Roos EM. Predictors of long-term effect from education and exercise in patients with knee and hip pain. Dan Med J 2014;61(7):A4867. Epub 2014/08/16. PubMed PMID: 25123117.
35. Thorstensson CA, Garellick G, Rystedt H, Dahlberg LE. Better management of patients with osteoarthritis: development and nationwide implementation of an evidence-based supported osteoarthritis self-management programme. Musculoskelet Care 2015;13(2):67–75, http://dx.doi.org/10.1002/msc.1085. Epub 2014/10/28. PubMed PMID: 25345913.