

Physiotherapy for osteoarthritis of the knee:
Predictors of outcome at one year

Cathy Chapple

A thesis for the degree of
Doctor of Philosophy
At the University of Otago, Dunedin,
New Zealand

26 September 2011

Abstract

Background: Knee osteoarthritis (OA) is a prevalent disease which contributes to significant pain and disability in older individuals. There is increasing evidence that physiotherapy, in the form of exercise or manual therapy, is an effective intervention. However, not all knee OA patients will benefit from physiotherapy treatment, and there are numerous other treatment options. Matching effective interventions to individuals who are most likely to benefit is desirable.

Aim: To identify predictors of successful response to physiotherapy treatment, evaluated at one year, for individuals with knee OA.

Methods: This observational cohort study was nested within a randomised controlled trial investigating effectiveness of physiotherapy for individuals with lower limb OA. Selection of variables for a standardised baseline assessment was informed by a systematic review identifying predictors of knee OA progression, and an inter-rater reliability study. A scale was developed to provide a valid and reliable method for assessing irritability, a potential predictor of outcome. Participants received physiotherapy including exercise and manual therapy. Outcome was evaluated at nine weeks and one year using the Western Ontario and McMaster osteoarthritis index (WOMAC) and global rating of change (GRC) transition scale; with response determined using OMERACT-OARSI (Clinical Trials Response Criteria Initiative and Outcome Measures in Rheumatology/ Osteoarthritis Research Society International) responder criteria. Nine week response was investigated as an additional predictor of successful outcome.

Data Analysis: Pre-test probability of success was the number of participants with a successful response following physiotherapy treatment, divided by the total number of participants in the treatment cohort.

Baseline variables with a univariate association with one year outcome ($p < 0.2$), were entered into multivariate logistic regression with backward elimination, to identify a set of variables ($p < 0.1$) that predicted successful outcome at one year. Accuracy statistics were used to calculate post-test probability of success for different numbers of retained predictors. A novel approach used post-estimation analysis to identify named combinations of variables, and associated probability of success.

Results: Pre-test probability of success was 35%. Six predictors of success were identified: posterior knee pain, disturbed sleep, absence of knee injury, instability, symptom duration (> 5 years), and female sex. The optimal model, with at least four out of six predictors, increased post-test probability of success to 66%. Presence of less than three out of six predictors decreased post-test probability of success to 11%. Using the post-estimation test, a five variable model (posterior knee pain, disturbed sleep, absence of knee injury, instability, and female sex), gave an 87% probability of success.

Data from the usual care control group did not fit the model, providing some evidence that variables were predictors of treatment response rather than natural progression of knee OA.

Nine week response gave an inferior prediction of probability of success at one year (52%), but combined with baseline predictors produced an optimal model of at least five out of seven predictors, with a post-test probability of 86%.

Conclusion: It is possible to use baseline variables, with or without nine week response, to predict physiotherapy treatment response at one year, for patients with knee OA. These findings represent the derivation stage of a clinical prediction rule which requires validation in future research.

Acknowledgements

My thanks and appreciation goes to my supervisory team, Professor Helen Nicholson, Professor G. David Baxter, and Dr J Haxby Abbott. I'm amazed and grateful for the time and effort you have put towards the production of this thesis.

Haxby, thanks for providing me with the opportunity and convincing me I could do it.

Helen, for always keeping your eye on the end game, and dragging me back on track whenever I veered off course, thank-you.

Dave, thanks for encouraging and supporting me throughout (was it carrot or was it stick?)

Special thanks to my two fellow PhD/MOA colleagues, Lexie and Dan, for generously sharing your time and knowledge. It was great to share the journey, and I couldn't have done it without you!

Thank-you Dr Jose Garcia, for revealing the mysteries of biomedical statistics, and helping me with the analysis.

Thanks Emma Tumilty for countless things, and for always doing them with a smile.

Thanks to Dr Marina Moss and the other MOA clinical trial administrators, Andrew Sullivan, Helen Ingrams, and Bob Matthews. Your professionalism and hard work was much appreciated, as well as your time and support for me.

Brigid Ryan, thanks for being a complete star!

Thank-you to everyone involved with the MOA trial: co-investigators, physiotherapists and all the participants. Without all your hard work and time I wouldn't have had a study! It is much appreciated.

Thanks also to Sinead Lodge and the staff in the School of Physiotherapy Clinics, for keeping me grounded in the real world of clinical practice, and for lots of words of encouragement and support. Special thanks to Sue Kennedy (my unofficial supervisor!).

Thanks to the Staff and postgraduate students at the School of Physiotherapy; so many of you have helped me at different times. It's been great to be part of the physiotherapy PhD community, there's nothing like going through it together, and I wish you all luck for completing your studies.

Thank-you to all my physiotherapy friends and colleagues, especially those from Dunedin Hospital. Thanks for your support and for believing that I could do it.

Thanks to all my friends in New Zealand and England for your help, encouragement and support; for not telling me too often I was mad, for excusing me from all sorts of things, for picking up the kids or keeping them occupied, and for all the numerous ways you have helped me to keep all the balls in the air!

Thanks especially to Mum and Dad and the rest of my family, for always encouraging me to pursue my goals, and having faith in me to achieve them.

A huge, massive, enormous thank-you to my two boys, Matt and Robbie, for sharing Mum with this PhD thing (even though it has "only got eight chapters"); for trying hard to let me study, for putting up with me not always being there, and for reminding me what's really important!

And finally, thank-you Mike for everything...

Table of Contents

1. Introduction	1
2. Systematic review of prognostic studies: variables that predict progression of knee osteoarthritis	11
2.1. Introduction	11
2.2. Methods	12
2.2.1. Search and identification of studies	12
2.2.2. Assessment of bias	15
2.2.3. Data extraction	18
2.2.4. Synthesis of evidence	18
2.3. Results	19
2.3.1. Search results	19
2.3.2. Study details	19
2.3.3. Assessment of Bias	20
2.3.4. Synthesis of evidence	23
2.3.5. Sensitivity analysis	34
2.4. Discussion	35
2.4.1. Comparison with previous research	36
2.4.2. Study limitations	38
2.4.3. Implications for future research	40
2.4.4. Summary	41
2.5. Implications for selection of baseline variables as potential predictors of treatment response for patients with knee OA	41
3. The concept of irritability and development of the i-score	43
3.1. Overview	43
3.2. Concept of irritability	43
3.3. Scale development and testing	45

3.3.1.	Item selection _____	45
3.3.2.	Scaling of responses _____	45
3.3.3.	Conversion of items to a scale _____	46
3.3.4.	Testing psychometric properties _____	46
3.4.	Methods _____	48
3.4.1.	Existing method for assessment of irritability _____	48
3.4.2.	Development of scale for assessment of irritability _____	48
3.4.3.	Testing psychometric properties of the i-score _____	50
3.4.4.	Recruitment _____	53
3.4.5.	Data analysis _____	54
3.5.	Results _____	55
3.6.	Discussion _____	64
3.6.1.	Previous research _____	64
3.6.2.	Strengths _____	64
3.6.3.	Limitations _____	66
3.6.4.	Future research _____	67
3.6.5.	Implications for use of the i-score _____	68
3.7.	Conclusions _____	69
4.	Inter rater reliability of baseline measures _____	70
4.1.	Introduction _____	70
4.2.	Methods _____	71
4.2.1.	Literature review _____	71
4.2.2.	Investigation of inter-rater reliability for clinical tests _____	72
4.2.3.	Data analysis _____	73
4.3.	Results _____	75
4.3.1.	Literature review _____	75
4.3.2.	Physical examination – investigation of inter-rater reliability of clinical measures _____	92
4.4.	Discussion _____	102

4.4.1.	Limitations and implications for future research	103
4.5.	Conclusion	106
5.	Study to identify predictors of outcome: Methods	107
5.1.	Chapter overview	107
5.2.	Study design	108
5.2.1.	Study setting	109
5.2.2.	Participants	110
5.2.3.	Assessors	113
5.3.	Baseline assessment	113
5.3.1.	Patient history	114
5.3.2.	Clinical features and physical examination	117
5.4.	Overview of trial procedure	124
5.4.1.	Randomisation	124
5.4.2.	Standardised treatment	124
5.4.3.	Follow-up assessments	125
5.5.	Outcome measures	126
5.6.	Response Criteria	127
5.7.	Data management and preparation	128
5.7.1.	Management of missing data	128
5.7.2.	Management of incomplete data due to surgical intervention	130
5.7.3.	Screening of baseline variables for analysis	131
5.7.4.	Cut-points in data	132
5.8.	Data analysis	133
5.8.1.	Overview of analysis	133
5.8.2.	Univariate analysis	135
5.8.3.	Multivariate analysis	135
5.8.4.	Confounders	136
5.8.5.	Combinations of predictor variables and post-test probability	136
5.8.6.	Secondary analyses	138

6.	Study to identify predictors of outcome: Results	142
6.1.	Overview	142
6.2.	Characteristics of knee OA cohort from the clinical trial	142
6.3.	Dropouts, missing data and surgical participants	147
6.4.	Baseline variables for prognostic study	149
6.4.1.	Screening of baseline variables	149
6.4.2.	Creation of cut-points in data	156
6.5.	Predictors of success at one year for participants with knee OA following physiotherapy intervention	157
6.5.1.	Successful outcome at one year	157
6.5.2.	Univariate analysis identifying baseline variables associated with success at one year	158
6.5.3.	Multivariate analysis identifying baseline variables associated with success at one year	159
6.5.4.	Stratified analysis	161
6.5.5.	Clusters of predictor variables, and post-test probability of success at one year	162
6.5.6.	Secondary analyses	170
6.5.7.	Sensitivity Analyses	181
7.	Study to identify predictors of outcome: discussion	197
7.1.	Chapter overview	197
7.2.	Predicting success with physiotherapy for patients with knee OA	197
7.2.1.	Posterior knee pain	198
7.2.2.	Absence of previous knee injury	200
7.2.3.	Disturbed sleep	202
7.2.4.	Instability	203
7.2.5.	Sex	204
7.2.6.	Symptom duration (> 5 years)	206
7.2.7.	Hip external rotator muscle strength	208

7.2.8.	Irritability_____	209
7.3.	Variables not identified as predictors _____	210
7.4.	Secondary analyses _____	214
7.4.1.	Do identified variables predict response to physiotherapy intervention or natural progression over one year? _____	214
7.4.2.	Nine week response as predictor of one year outcome_____	215
7.4.3.	Predictors of poor outcome at one year_____	218
7.5.	Sensitivity analyses _____	218
7.6.	Comparison with previous findings _____	220
7.7.	Strengths and limitations _____	221
7.7.1.	Strengths _____	221
7.7.2.	Limitations_____	223
7.8.	Summary _____	228
7.8.1.	Clinical implications _____	228
7.8.2.	Research implications _____	230
8.	Predictors of outcome following physiotherapy for individuals with knee osteoarthritis: Conclusions _____	234
8.1.	Summary of major findings of prognostic study_____	234
8.2.	Summary of major findings of supplementary studies _____	237
8.2.1.	Systematic review identifying predictors of progression in patients with knee OA_____	237
8.2.2.	The concept of irritability and development of the i-score_____	238
8.2.3.	Inter-rater reliability of clinical tests used to evaluate patients with knee OA _____	239
8.3.	Strengths and limitations _____	240
8.4.	Research implications _____	241
8.5.	Clinical implications _____	243

8.6. Conclusion	244
-----------------	-----

References	247
------------	-----

Appendices – available on CD-ROM (attached)

- Appendix A – Protocol for conducting systematic review
- Appendix B – Search terms for Medline (adapted for other databases)
- Appendix C – Operational definitions for assessment of bias tool
- Appendix D – Data extraction form for systematic review
- Appendix E – Details of studies included in systematic review
- Appendix F – Tables B and C, results from all studies in systematic review
- Appendix G – Data for inter-rater reliability of the i-score
- Appendix H – Baseline assessment form
- Appendix I – Ethics approval form
- Appendix J – Inter rater reliability assessment form
- Appendix K – Manual of operating procedures for baseline assessment
- Appendix L – Results of inter rater reliability for all clinical tests of the hip and knee
- Appendix M – Patient Global Rating of Change (GRC) transition scale
- Appendix N – Standardised physiotherapy treatment protocols
- Appendix O – Baseline and one year data for WOMAC and GRC
- Appendix P – OARSI responder data

List of Tables

Table 2.1 Scoring tool for assessment of bias, for articles included in the systematic review.....	17
Table 2.2 Assessment of bias scores for included papers in the systematic review.....	21
Table 2.3 Evidence from high quality studies for predictors of progression of knee osteoarthritis: demographics and patient history (self-report measures).....	25
Table 2.4 Evidence from high quality studies for predictors of progression of knee osteoarthritis: physical examination and clinical tests.....	29
Table 2.5 Comparison of findings from systematic reviews that identify variables with strong evidence for progression of knee osteoarthritis.....	37
Table 3.1 The i-score: a scale for assessment and reporting of irritability for patients with knee or hip osteoarthritis.....	56
Table 3.2 Descriptive statistics for the i-score applied to all knee osteoarthritis participants.....	57
Table 3.3 Operational definitions: aggravating activities for knee and hip osteoarthritis.....	58
Table 3.4 Results of Nominal Group Technique to decide important factors for assessing irritability for knee and hip osteoarthritis.....	59
Table 3.5 Descriptive statistics for the i-score applied to participants in the treatment arms of a clinical trial.....	60
Table 3.6 Psychometric properties of the i-score.....	63
Table 4.1 Inter-rater reliability of patient history variables from review of literature.....	76
Table 4.2 Inter-rater reliability of physical examination variables from literature review.....	81

Table 4.3 Inter-rater reliability of clinical measures reported with ICC (2,1) in participants with knee or hip osteoarthritis.....	93
Table 4.4 Inter-rater reliability of clinical variables reported with kappa (κ) in participants with knee or hip osteoarthritis.....	95
Table 5.1 Inclusion/exclusion criteria for the MOA clinical trial.....	111
Table 5.2 OMERACT-OARSI responder criteria to define outcome at one year for participants with knee osteoarthritis following physiotherapy intervention.....	127
Table 5.3 Magnitude of likelihood ratios (LR) and effect on post-test probability.....	134
Table 5.4 OMERACT-OARSI responder criteria to determine poor outcome at one year for participants with knee osteoarthritis following physiotherapy intervention.....	140
Table 6.1 Characteristics of participants with knee osteoarthritis enrolled in the clinical trial.....	144
Table 6.2 Baseline and one year WOMAC pain and function subscale scores for treatment and usual care groups.....	146
Table 6.3 Characteristics of completers and dropouts.....	148
Table 6.4 Baseline variables included in univariate analyses: inter-rater reliability.....	150
Table 6.5 Reasons for excluding variables from analysis.....	155
Table 6.6 Responders at one year and probabilities of successful outcome.....	158
Table 6.7 Baseline variables associated with success at one year for participants with knee osteoarthritis following physiotherapy intervention: univariate analysis.....	159
Table 6.8 Predictors of success at one year for participants with knee osteoarthritis following physiotherapy intervention: multivariate analysis.....	160

Table 6.9 Predictors of success at one year stratified by sex.....	162
Table 6.10 Different test conditions for baseline predictors with associated likelihood ratios and post-test probabilities.....	164
Table 6.11 Different test conditions for baseline predictors with associated likelihood ratios and post-test probabilities stratified by sex.....	166
Table 6.12 Model variants showing specific combinations of named predictors and range of probability for success at one year.....	169
Table 6.13 Predictors of success at one year when treatment (yes/no) forced into multivariate analysis.....	170
Table 6.14 Different test conditions for baseline predictors with associated likelihood ratios and post-test probabilities using data from usual care group.....	172
Table 6.15 Success at nine weeks for participants with knee osteoarthritis following physiotherapy intervention.....	173
Table 6.16 Odds of success at one year following success at nine weeks: univariate analysis.....	174
Table 6.17 Nine week response as a predictor of success at one year.....	175
Table 6.18 Probability of success at one year with baseline predictors <i>and</i> success at nine weeks.....	176
Table 6.19 Nine week response and named predictors: probability of success at one year.....	178
Table 6.20 Number of responders and probability of poor outcome at one year.....	179
Table 6.21 Baseline variables associated with poor outcome at one year: univariate analysis.....	180
Table 6.22 Univariate associations of BMI and age with success at one year, displaying different data formats.....	182

Table 6.23 Predictors of success at one year: multivariate sensitivity analysis for age (> 57 years).....	183
Table 6.24 Post-test probability of success: sensitivity analysis for age (> 57 years).....	184
Table 6.25 Baseline predictors associated with success at one year in the modified cohort: univariate analysis.....	186
Table 6.26 Baseline predictors of successful outcome at one year in modified cohort: multivariate sensitivity analysis.....	187
Table 6.27 Baseline predictors of successful outcome at one year in modified cohort: multivariate sensitivity analysis stratified by sex.....	188
Table 6.28 Post-test probability for females: sensitivity analysis with modified cohort.....	189
Table 6.29 Odds of success at one year following success at nine weeks in the modified cohort: sensitivity analyses.....	190
Table 6.30 Nine week response as predictor of success at one year in modified cohort: sensitivity analysis.....	191
Table 6.31 Summary of main results.....	193

List of figures

Figure 1 Design of research programme to show how supplementary studies contribute to the main prognostic study.....	10
Figure 2.1 Flow diagram showing the different phases of study retrieval for the systematic review.....	15
Figure 3.1 Procedure for nominal group technique.....	51
Figure 3.2 Histogram showing the distribution of the i-score in knee osteoarthritis participants.....	57
Figure 3.3 Distribution of pain scores from physical performance measures and change in pain (NPRS) over 24-hours.....	61
Figure 4.1 Bland Altman plot showing agreement between two raters for hip flexion range of motion (in degrees).....	97
Figure 4.2 Bland Altman plot showing agreement between two raters for assessment of quadriceps strength “knee extension at ninety degrees”.....	98
Figure 4.3 Bland Altman plot showing agreement between two raters for assessment of hamstrings strength “knee flexion at 0-20 degrees”.....	99
Figure 4.4 Bland Altman plot showing agreement between two raters for hip internal rotation muscle strength.....	99
Figure 4.5 Bland Altman plot showing agreement between two raters for hip external rotation muscle strength.....	100
Figure 4.6 Bland Altman plot showing agreement between two raters for hip flexion muscle strength.....	100
Figure 4.7 Bland Altman plot showing agreement between two raters for hip abduction muscle strength.....	101
Figure 4.8 Bland Altman plot showing agreement between two raters for forty metre self-paced walk.....	102
Figure 5.0 Overview of analysis (with reference to results section).....	141

Figure 6.1 Flow chart showing study recruitment and retention of participants with knee OA in the clinical trial.....	143
Figure 6.2 Distribution of baseline WOMAC scores for participants with knee OA randomised to the treatment cohort in the clinical trial.....	145

Publications

Peer reviewed journal

Chapple, C.M., Nicholson, H., Baxter, G.D., Abbott, J.H. 2011. Patient characteristics that predict progression of knee osteoarthritis: A systematic review of prognostic studies. *Arthritis Care and Research*, 63, 1115-25.

The candidate conceived and designed the study; acquired, analysed and interpreted the data; was primary author, drafted and revised the manuscript.

Peer reviewed conference presentations

Chapple, C.M., Baxter, G.D., Nicholson, H., Abbott, J.H. 2011. Will this patient with knee osteoarthritis benefit from physiotherapy. NZMPA Scientific Conference, Rotorua, New Zealand. 27-28 August 2011

The candidate designed the study; acquired, analysed and interpreted the data; composed and gave the presentation.

Chapple, C.M., Baxter, G.D., Nicholson, H., Abbott, J.H. 2011. How hard can I push? Investigating the concept and assessment of irritability. NZMPA Scientific Conference, Rotorua, New Zealand. 27-28 August 2011

The candidate conceived and designed the study; acquired, analysed and interpreted the data; composed and gave the presentation.

Abbott JH, Chapple C, Pinto D, Wright AA. 2011. Knee changes from the sports injury to the TKA, Part 3: The treatment: Management of OA using exercise therapy, manual therapy, or both. Part 3 of 4 sponsored by the Sports, Research, Orthopaedics, and Geriatrics Sections, Combined Sections Meeting of the American Physical Therapy Association, New Orleans, LA, USA. February 9-12, 2011.

The candidate was responsible for composing and presenting section on "Exercise therapy for hip and knee OA"; designed the study, acquired, analysed, interpreted data, composed and gave presentation "Patient prognosis: knee OA";

for main MOA study designed knee assessment and treatment protocols, acquired and managed knee data, analysis of outcome for knee participants.

Chapple, C.M., Nicholson, H., Baxter, G.D., Abbott, J.H. 2009. How hard can I push? Investigating the concept and assessment of irritability. *New Zealand Journal of Physiotherapy*, 37, 153. Presented at the Southern Physiotherapy Symposium 5, Queenstown, New Zealand. 20-22 November 2009.

The candidate conceived and designed the study; acquired, analysed and interpreted the data; composed and gave the presentation.

List of abbreviations

ACR	American College of Rheumatology
AIC	Akaike Information Criterion
AMED	Allied and complementary MEDicine
AP	Antero-Posterior
AUC	Area Under the Curve
BIC	Bayesian Information Criterion
BMI	Body Mass Index
CI	Confidence Interval
CINAHL	Cumulative Index to Nursing and Allied Health Literature
COG	Centre Of Gravity
CPR	Clinical Prediction Rule
CTT	Classical Test Theory
FWB	Full Weight Bearing
GOF	Goodness of Fit
GP	General Practitioner
GRC	Global Rating of Change
HHD	Hand Held Dynamometer
HR	Hazard Ratio
ICC	Intraclass Correlation Coefficient
IRT	Item Response Theory
ITT	Intention To Treat
JSN	Joint Space Narrowing
JSW	Joint Space Width
KL	Kellgren Lawrence (grade)
KOS-ADLS	Knee Outcome Survey-Activities of Daily Living Scale
LAA	Longitudinal Arch Angle
LBP	Low Back Pain
LOCF	Last Observation Carried Forward
LR	Likelihood Ratio
MCID	Minimal Clinically Important Difference
MI	Multiple Imputations
MRI	Magnetic Resonance Imaging
MOA	Management of Osteoarthritis
MOP	Manual of Operating Procedures
NGT	Nominal Group Technique
NPRS	Numeric Pain Rating Scale
NWB	Non Weight Bearing

OA	Osteoarthritis
OARSI	Osteoarthritis Research Society International
OMERACT	Clinical Trials Response Criteria Initiative and Outcome Measures in Rheumatology
OR	Odds Ratio
PA	Postero-Anterior
PASS	Patient Acceptable Symptom State
PBSI	Pain Belief Screening Instrument
PICOS	Population, Interventions, Comparators, Outcomes and Study designs
PPM	Physical Performance Measure
PRISMA	Preferred Reporting Items for Systematic reviews and Meta-Analyses
PWB	Partial Weight Bearing
RCT	Randomised Controlled Trial
ROC	Receiver Operating Characteristic
ROM	Range Of Motion
RPE	Rate of Perceived Exertion
RR	Relative Risk ratio
SD	Standard Deviation
SEM	Standard Error of Measurement
SLR	Straight Leg Raise
SPW	Self Paced Walk (40 meters)
SST	Sit to Stand Test (30 sec)
TKR	Total Knee Replacement
TUG	Timed Up and Go
VAS	Visual Analogue Scale
WHO	World Health Organisation
WOMAC	Western Ontario and McMaster Universities osteoarthritis index

Symbols

α	Cronbach's alpha
κ	Kappa statistic
p_0	minimal accepted level of reliability
p_1	desirable level of reliability
95%CI	95% confidence interval

1. Introduction

Physiotherapy can provide effective interventions for individuals with knee osteoarthritis (OA), but not all individuals with knee OA will benefit from physiotherapy. There is strong evidence that exercise therapy in particular produces good outcomes for patients with knee OA (Roddy et al., 2005; National Collaborating Centre for Chronic Conditions, 2008; Zhang et al., 2008; Fransen and McConnell, 2009), while there is growing evidence that adding manual therapy is also beneficial (Deyle et al., 2000, 2005; Jansen et al., 2011). However, effect sizes from pooled data are small to moderate (Fitzgerald and Oatis, 2004; Fransen and McConnell, 2009; Jansen et al., 2011). Modest effect sizes may be due to sub-groups within study samples responding differently to treatment (Sharma, 2003; Fitzgerald and Oatis, 2004). For example, knee joint alignment or laxity in patients with knee OA may alter response to exercise programmes, suggesting treatment should be tailored to individuals or sub-groups (Sharma, 2003). It has been argued that identifying patients more likely to respond to specific interventions would improve treatment outcomes (Hancock et al., 2009).

Osteoarthritis of the knee is a common disease with a prevalence of 12.5% in populations over forty five years of age (Zhang et al., 2010). As OA tends to increase with age and there is demographic ageing of the population in New Zealand and many other Western countries, there is a projected increase in prevalence (Felson, 1995; Access Economics, 2005, 2010). Furthermore, the lifetime risk of developing symptomatic knee OA is estimated as 44.7 % (Murphy et al., 2008). The disease affects quality of life and contributes to disability, as well as being a considerable and growing economic burden to society (Peat et al., 2001; Access Economics, 2005, 2010). The size of the problem heightens the importance

of developing and utilising effective treatments and management strategies for individuals suffering with knee OA. The problem is further exacerbated by the heterogeneity of the disease resulting in a wide variety of clinical presentations (Hunter, 2009; Zhang et al., 2010).

These issues were encapsulated by Fransen (2004, p.485) in the following quotation:

“If certain patient characteristics could identify either responders or non-responders to physiotherapy or graded exercise, valuable evidence-based referral guidelines could be established. Much wasted effort could be avoided and physiotherapy treatment might become more accessible to those patients most likely to benefit.”

Despite recognition of the need to identify sub-groups of responders to treatment within the population of knee OA patients, there has been limited research into identifying predictors of outcome. One study identified female sex, absence of depressive symptoms, history of the use of complementary medicine and low comorbidities as predictors of response to a three to four week in-patient rehabilitation programme for patients with knee and hip OA (Weigl et al., 2006). However, this study limited potential predictors to sociodemographic variables, lifestyle risk factors and psychological status of the patients, and did not include any clinical tests or physical measures. Generalizability is also limited by the mode of treatment delivery being on an in-patient basis, whereas New Zealand and many other countries tend to provide similar physical rehabilitation on an out-patient basis. There remains therefore the need to identify baseline predictors of response to a physiotherapy treatment programme delivered in an out-patient setting. Additionally, as knee OA is a slowly progressing disease (Felson, 1995), long-term follow-up of outcome should be considered, with one year follow-up arguably more relevant than the six month follow-up used in the previous study (Weigl et al., 2006).

An earlier study included an evaluation of baseline predictors of responsiveness to an out-patient physiotherapy programme (Fransen et al., 2001). In this RCT severe loss of medial joint space (measured radiographically) reduced responsiveness to physiotherapy treatment. The only other variables considered were BMI, age, and symptom duration, leaving many potential predictors of treatment response yet to be investigated. Additionally outcome was evaluated at sixteen weeks, which similar to the aforementioned study may be of limited relevance to a chronic, slowly progressive disease such as knee OA.

Highlighting gaps in current knowledge led to development of the main aim of the thesis: to identify predictors of successful outcome at one year following physiotherapy intervention for patients with knee OA.

Development of a clinical prediction rule (CPR) offered a credible approach to answering the main aim of the thesis. CPRs combine assessment variables from patient history and physical examination to produce clusters which describe numeric probability of an outcome (Beattie and Nelson, 2006; Stanton et al., 2010). That outcome may be diagnosis, prognosis or, frequently in physiotherapy practice, treatment response (Fritz, 2009). There is a well-established three-step method for developing a CPR involving derivation of the rule, validation or testing it, and assessment of its impact on clinical practice (McGinn et al., 2000). This thesis will describe step one in that process.

While development of CPRs has become increasingly popular in physiotherapy research in the last decade, caution has been advised (Fritz, 2009; Stanton et al., 2010). Firstly, a problem should warrant the time and expense required to develop a rule: specifically when a condition is heterogeneous and therefore likely to produce subgroups of patients with different responses to treatment (Beattie and Nelson, 2006; Fritz, 2009). As noted above knee OA meets this requirement. Secondly development is indicated when there are different

treatment options available. CPRs will assist clinicians to choose between them, selecting treatments with the greatest probability of success (McGinn et al., 2000; Fritz, 2009). Patients with knee OA have the option of numerous conservative methods of management for knee OA, in addition to major joint surgery. In the OARSI guidelines, there are eleven recommendations about non-pharmacological modalities; a further eight addressing pharmacological modalities, and five recommendations about surgical options (Zhang et al., 2008). Clinical decision-making about the suitability of a particular intervention for an individual patient is a complex process. Any intervention is associated with risks and costs which must be weighed against likely benefit.

Surgical interventions for knee OA, such as total joint replacement, whilst effective, are associated with significant costs and are not suitable or available for all patients (Zhang et al., 2008). Pharmacological interventions may have adverse consequences and demonstrate limited effectiveness in modifying the disease process (Hunter, 2009). However exercise and manual therapy treatment provided by physiotherapists is effective for reducing pain and improving physical function in patients with knee OA (Deyle et al., 2000, 2005; National Collaborating Centre for Chronic Conditions, 2008; Fransen and McConnell, 2009). It is also a low risk intervention, involving lower costs than surgery or long-term use of some pharmaceuticals (Beattie and Nelson, 2008), making it an attractive option for management of knee OA in some individuals. Therefore the heterogeneity of knee OA and the range of treatment options provide good justification for embarking on development of a CPR to determine those patients most likely to benefit from physiotherapy treatment.

Another factor that must be considered is whether identified variables are predictors of treatment response, or are instead predictors of the natural course of the disease regardless of treatment intervention. As knee OA is characterised by

periods of disease activity and decline, interspersed with periods of relative inactivity and even reduction of symptoms (National Collaborating Centre for Chronic Conditions, 2008, Bijlsma et al., 2011), this distinction is particularly important. This point has been highlighted in recent critiques of clinical prediction rules, with the recommendation that predictors of response to treatment can only be identified with any assurance if compared to a control group in a randomised controlled trial (Stanton et al., 2010). However, it may still be acceptable to use a single arm study as a preliminary step in CPR development, to reduce the number of potential predictors of treatment outcome prior to testing in a subsequent randomised controlled trial (RCT) (Fitzgerald, 2010). The current prognostic study was nested within an RCT in which one group did not receive physiotherapy treatment (Abbott et al., 2009). There was therefore potential to compare predictors of treatment response with predictors of natural course of disease over one year. In addition patient characteristics associated with outcome in the form of disease progression were identified by conducting a systematic review (Chapter 2). These findings were used to inform the choice of baseline variables in the main study.

Patients with knee OA may opt for, or be referred to, physiotherapy for a variety of reasons including personal preference, doctor preference, presence of comorbidities limiting other treatment options or rationed access to specialist clinics or surgery (Ballantyne et al., 2007; Porcheret et al., 2007; Thorstensson et al., 2009b; Cottrell et al., 2010). Frequently physiotherapists are obligated to provide treatment and may not have the option to deny treatment based on presence or absence of predictive variables. Therefore evaluation of treatment response provides additional information, which could contribute to prediction of long-term outcome. A secondary analysis was performed to determine whether

immediate response on completion of the treatment programme was a predictor of successful outcome at one year.

In many clinical trials the focus is on identifying improvement in patient status as a response to intervention (Tubach et al., 2005; Kvien et al., 2007). However, it can be argued that identifying participants at risk of worsening as a result of intervention is also clinically relevant. There are numerous types of treatment available for patients with knee OA (Zhang et al., 2008), so knowledge that participation in physiotherapy treatment is likely to result in poor outcome in some cases would influence both patient and clinician to try another option. The possibility of identifying baseline variables that would predict poor outcome at one year following physiotherapy intervention was an additional research question.

Irritability describes the ease with which a condition may be exacerbated by movement, and is an aspect of symptom behaviour with the capacity to influence patient outcomes (Maitland, 1991). Although it has not previously been investigated in patients with knee OA, it was recognised as a potential predictor of outcome with the hypothesis that lower irritability could be a predictor of success at one year following physiotherapy intervention.

Irritability is a concept familiar to many physiotherapists. However, closer examination revealed a lack of standardised assessment and reporting, which has implications for clinical practice as well as creating a problem in the research setting. Therefore an essential pre-requisite to investigating irritability as a potential predictor of outcome was to develop an assessment tool with good psychometric properties. The investigation of irritability, and development and testing of an irritability scale is reported in Chapter 3.

Inter-rater reliability of clinical variables is frequently reported in studies identifying predictors of outcome or CPR development. Good inter-rater

reliability indicates repeated use of the measure is likely to produce consistent results, free from error, and representing the true score (Portney and Watkins, 2000). It allows confident use of a measure in larger populations by different testers or clinicians (Portney and Watkins, 2000). This is important when test results are used to guide clinical decisions (Hicks et al., 2003). Inter-rater reliability was an important consideration in selecting potential predictor variables to include in the baseline assessment, and formed the basis of a research question. Evidence of inter-rater reliability of some clinical variables was found in the literature. Where evidence was lacking in either an older population or a knee OA population, clinical variables were the subject of an inter-rater reliability investigation. The reliability study is reported in Chapter 4.

A guiding principle of the research programme was to select variables that could be easily measured in usual clinical environments. In this setting assessment techniques should be quick and easy to perform and not involve sophisticated equipment or elaborate methods of analysis. These considerations make the findings of the study more generalisable and avoid introducing unnecessary barriers to implementation of findings. A similar consideration was to avoid making inclusion/exclusion criteria too restrictive, with the intention of making the trial population as reflective as possible of usual clinical practice.

Prognostic information is considered valuable by patients, but not always included in initial consultation for musculoskeletal pain (Mallen and Peat, 2009). It can be used to educate patients enabling them to make informed choices about future care (Mallen and Peat, 2009). Health practitioners use prognostic information to identify risk factors for disease and to guide clinical decision-making, including directing patients to treatments with the greatest probability of success (Beattie and Nelson, 2006; Childs and Cleland, 2006). This limits

unnecessary financial expenditure, in addition to reducing time and effort wasted on ineffective interventions.

The programme of research for the thesis was conducted within an RCT, the Management of Osteoarthritis (MOA) clinical trial, which investigated the clinical effectiveness and cost-effectiveness of physiotherapy treatment for knee and hip OA (Abbott et al., 2009). It was a 2X2 factorial randomised controlled trial, conducted over twenty-four months in Dunedin, New Zealand. Recruitment of participants took one year. Seven treatment sessions were provided over a nine week period, with two booster sessions at sixteen weeks. Follow-up assessments of outcome were conducted at nine weeks, twenty-six weeks and one year. Standardised interventions included manual therapy, exercise therapy, a combination of both, or usual care. Further details are available in the trial protocol (Abbott et al., 2009), with points relevant to the conduct of the thesis research outlined in the relevant chapters.

In summary the main aim of the thesis is:

- To identify predictors of successful response at one year for individuals with knee OA following physiotherapy treatment.

The hypotheses examined in this thesis are that baseline variables can be identified as predictors of successful outcome at one year following physiotherapy intervention for individuals with knee OA; secondly, probability of success at one year can be predicted by presence or absence of clusters of predictors (models).

The supplementary studies of irritability and reliability addressed secondary aims:

- To develop a valid and reliable scoring tool for the assessment of irritability.
- To establish inter-rater reliability of tests and measures used in the standardized clinical assessment of knee OA participants.

Additional research questions outlined above considered whether response was due to treatment intervention or was natural progression over one year; if response at completion of the treatment programme predicted outcome at one year; and whether baseline variables could predict worsening of knee OA at one year following physiotherapy intervention. These questions were addressed in secondary analyses.

These aims and questions have produced a programme of research which is outlined in Figure 1.

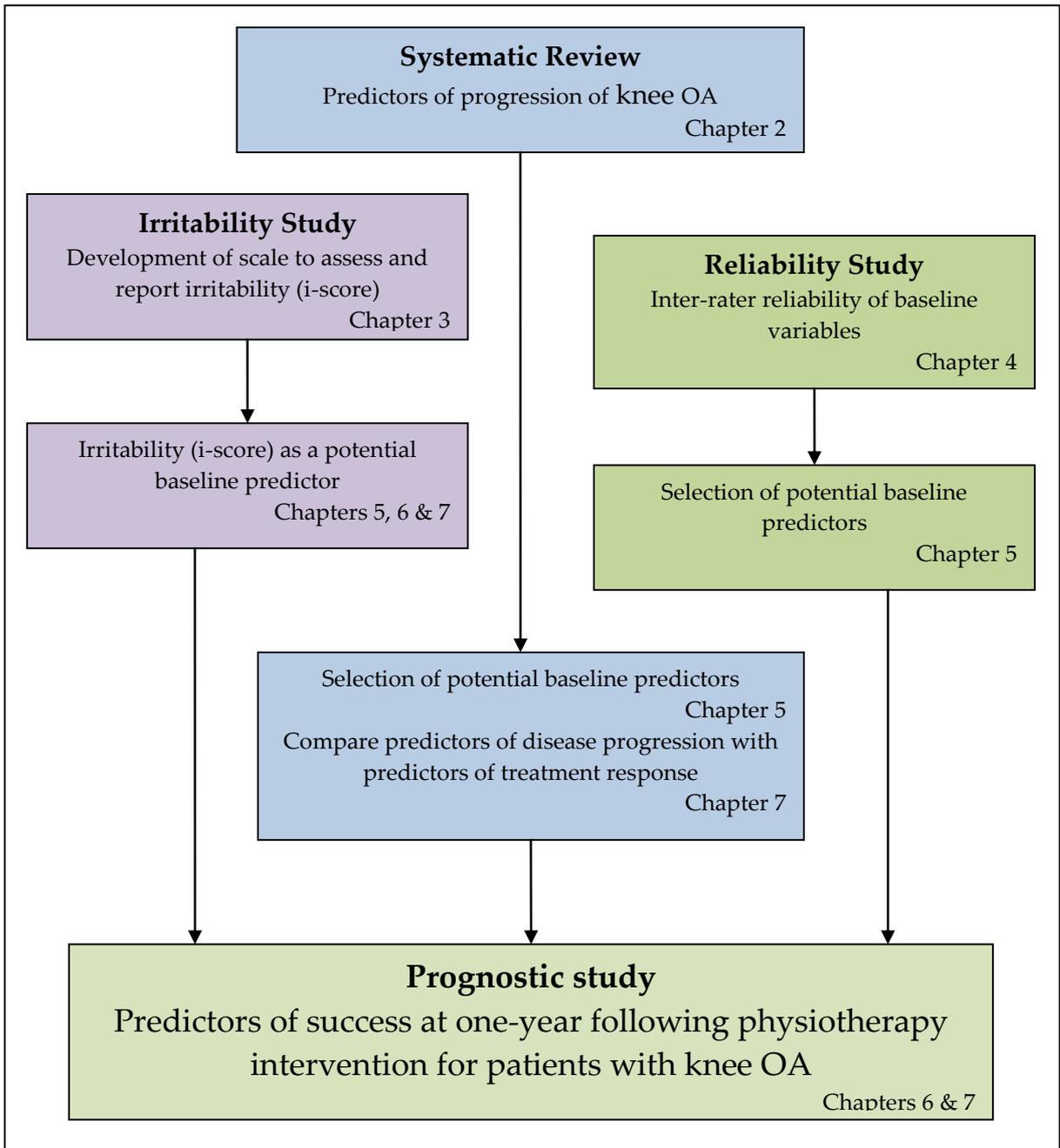


Figure 1 Design of research programme to show how supplementary studies contribute to the main prognostic study

2. Systematic review of prognostic studies: variables that predict progression of knee osteoarthritis

2.1. Introduction

Increased interest and research regarding predictors of outcome with physiotherapy intervention has occurred in recent years (Fritz, 2009; Stanton et al., 2010), although few studies have sought to identify predictors of outcome following physiotherapy intervention for knee osteoarthritis (OA). Conversely there are many studies that have investigated natural progression of the disease and associated predictive variables. This latter body of literature will be useful for the current study, identifying potential baseline variables for inclusion in the main study, that have existing evidence of association with outcome. Furthermore, it will be possible to compare and contrast predictors of outcome, with and without physiotherapy intervention.

Systematic review was selected as the method for reviewing the literature on predictors of knee OA progression. Systematic reviews involve “comprehensive collection, summary and often analysis of all studies relevant to a focussed research question” (p.87 Simunovic et al., 2009). A fundamental aim is to minimise bias by having clearly defined processes for finding and screening articles; assessing the quality of included articles; retrieving and analysing data; and summarizing findings.

Studies investigating predictors of knee OA progression are predominantly observational cohort studies. Specific issues relevant to systematic reviews of observational studies have been identified, such as selection bias, information bias and confounding of reviewed studies (Simunovic et al., 2009). Associated recommendations have been used to guide the conduct of this systematic review

(Centre for Reviews and Dissemination, 2008; Manchikanti et al., 2009; Simunovic et al., 2009).

Previous systematic reviews have sought to identify predictors of knee OA progression using radiographic change (Belo et al., 2007; Tanamas et al., 2009), or functional decline and change in pain (van Dijk et al., 2006) as outcome measures. Usual clinical practice is to gather information from both radiographs *and* patient reported symptoms. Therefore to enhance clinical relevance, the approach of this review was to include any validated clinical information for both potential predictor variables and measures of outcome.

The aim of this systematic review is to identify patient characteristics that can be used by health care practitioners to predict progression of knee OA.

2.2. Methods

A protocol was designed for the conduct of the review with reference to the preferred reporting items for systematic reviews and meta-analyses (PRISMA) guidelines (Moher et al., 2009) (Appendix A). The PRISMA statement acknowledges that a systematic review is an iterative process and that modifications may need to be made to study protocols (Moher et al., 2009). This is acceptable practice if changes are recorded and justified.

2.2.1. Search and identification of studies

A comprehensive search strategy was developed to detect prognostic studies, with an approach using a framework including population, interventions, comparators, outcomes and study designs (PICOS) guiding selection of terms (Wilczynski et al., 2004; Wilczynski and Haynes, 2005; Furlan et al., 2006). A search strategy was developed for Medline and adapted for Embase, the Cumulative index to Nursing and Allied Health Literature (CINAHL), Allied and Complementary Medicine (AMED), Web of Science, and Web of Knowledge (Appendix B). Reference lists of identified reviews were mined for further studies

(van Dijk et al., 2006; Belo et al., 2007; Tanamas et al., 2009). Neither study type nor language was limited (Furlan et al., 2006; Simunovic et al., 2009). An initial search was conducted in August 2009 and updated in November 2010.

Inclusion/exclusion criteria were determined *a priori*. Study participants were adults (age > eighteen years, any sex, any duration of symptoms), with knee OA classified by clinical or radiographic reference standards (Zhang et al., 2010). Cohort studies had to describe any intended exposure; included randomised controlled trials (RCT) had to describe their interventions. Comparison was between progression and non-progression of knee OA. Outcome was progression of disease defined as: change in functional status or pain report as evaluated with recognised instruments such as the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) or the visual analogue scale (VAS); or radiographic change defined as an increase in Kellgren Lawrence (KL) grade or joint space narrowing (JSN) score, increase in osteophytes or decrease in joint space width (JSW). As study design was not limited, RCTs with potential for inclusion had to demonstrate a non-significant effect of intervention or had to report the placebo group separately. All studies had to identify prognostic factors at baseline, and report a statistical association (or lack of association) with outcome. To ensure clinical relevance, baseline variables had to include patient characteristics such as pain or functional limitation, psychosocial factors, demographic information, body mass index (BMI), dietary intake, or gait abnormalities. All measures had to be in routine clinical use and not require sophisticated equipment or complex analysis. Minimum time for follow up was set at one year to allow sufficient time for progression to occur and to match the outcome period of the prognostic study.

Studies were excluded if participants had co-morbidities such as rheumatoid arthritis, cancer, osteoporosis, joint infection; generalised OA where knee OA results were not reported separately; or previous surgery for knee OA.

Duplicate reports from the same study reporting the same data were excluded. To ensure applicability of review findings to routine clinical practice, any study that reported exclusively on radiographic variables or laboratory tests with no reference to patient presentation was excluded.

The search strategy and inclusion/exclusion criteria were piloted to ensure good coverage of the field, key articles were identified and retained, and irrelevant articles discarded.

Screening of articles using inclusion/exclusion criteria was completed independently by two reviewers, initially focussing on titles and abstracts, and then full text versions of retrieved articles. Reasons for exclusion were documented. Consensus on final inclusion of articles was reached following discussion; a third reviewer was available to resolve any outstanding disagreements. The process is summarised in Figure 2.1.

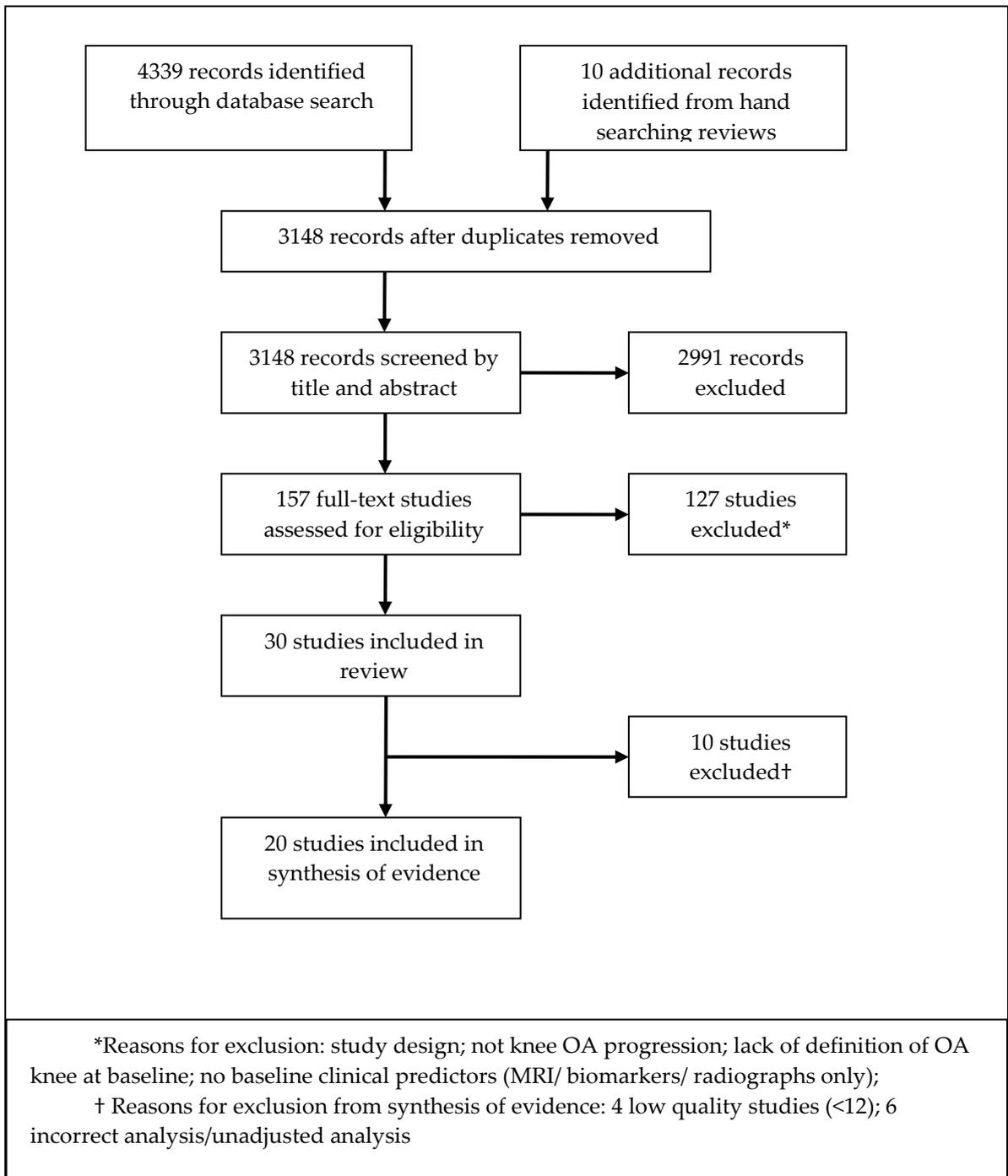


Figure 2.1 Flow diagram showing the different phases of study retrieval for the systematic review

2.2.2. Assessment of bias

Selected articles were assessed for bias using a scoring tool (Table 2.1) adapted from Wright et al. (2009), and based upon a similar tool used in previous

studies of musculoskeletal problems (Lievence et al., 2002; van Dijk et al., 2006; Belo et al., 2007; Tanamas et al., 2009). The tool was specifically aimed at defining study quality in terms of how potential sources of bias were addressed, rather than solely on quality of reporting. The revised tool addressed six potential sources of bias in prognostic studies: study participation (items A – D); study attrition (items E – G); prognostic factor measurement (items H, J – L); outcome measurement (items M – O); confounding measurement and statistical adjustment (multiple items, I and S); reporting and analysis of results and conclusions (items P – T) (Hayden et al., 2006; Centre for Reviews and Dissemination, 2008). Confounding and reporting of conclusions are considered particularly relevant in observational studies (Hayden 2006). Detailed operational definitions for the assessment of bias tool are included in Appendix C. For each criterion, articles scored one if they met the operational definition or zero if they did not meet it; if it was not applicable; or there was insufficient information to make a decision. The maximum score was twenty. Similar to other studies, an arbitrary cut off of $\geq 60\%$ of criteria met (i.e. a raw score of \geq twelve) defined studies of high quality (Lievence et al., 2002; van Dijk et al., 2006; Belo et al., 2007).

Table 2.1 Scoring tool for assessment of bias, for articles included in the systematic review

Study Participation	
A	Inception cohort (early/uniform point of disease)
B	Description of study population
C	Description of inclusion/exclusion criteria
D	Response $\geq 75\%$ for cohorts and controls
Study Attrition	
E	Follow-up of at least 12 months
F	Dropouts/Loss to follow-up $\leq 20\%$
G	Information completers vs. loss to follow-up
Measurement and data presentation	
H	Prospective data collection
I	Treatments/exposure of cohort fully described and standardised
J	Clinically relevant prognostic factors
K	Standardised or valid measures of prognostic factors
L	Data presentation of the most important prognostic factors
M	Clinically relevant outcome measures
N	Standardised or valid outcome measures
O	Data presentation of the most important outcome measures
Analysis and presentation of results	
P	Appropriate analysis techniques (provision of OR, RR, HR)
Q	Prognostic model is presented (multivariate techniques)
R	Sufficient numbers (10 subjects per variable)
S	Confounders identified and accounted for
T	Conclusions drawn accurately from results
OR = odds ratio; RR = relative risk; HR = hazard ratio	

Two reviewers, who were un-blinded to authors, used the tool to independently score included articles. Level of agreement beyond chance was reported using the kappa statistic (κ). Discussion was used to reach consensus on any scoring disagreements; a third reviewer was available to resolve any lack of agreement.

2.2.3. Data extraction

Study details and results were extracted from included articles by two unblinded reviewers. A form to guide and document a systematic data extraction process was developed and tested (Appendix D). Results included odds ratios (OR), relative risk (RR), hazard ratios (HR) or p-values. Adjusted results were extracted, where possible, to address the problem of confounding (Simunovic et al., 2009). Sources of funding were noted.

2.2.4. Synthesis of evidence

Options for synthesis of evidence included meta-analysis. However, heterogeneity is a common problem in prognostic studies and is an argument against pooling data for a formal meta-analysis (Altman, 2001; Centre for Reviews and Dissemination, 2008). An alternative was considered based upon three well-established domains of quality, quantity and consistency of findings from included studies (West et al., 2002). Studies had to be of high quality, scoring $\geq 60\%$ (≥ 12) with the assessment of bias tool. Additionally, univariate analyses adjusted for confounders, multivariate analyses or stratified analyses were considered as appropriate analyses for identification of prognostic factors (Centre for Reviews and Dissemination, 2008). Quantity of findings was addressed by reporting the number of studies investigating a particular variable; combined sample size; and magnitude of effect. Consistency was the extent to which similar findings were found in studies of similar design (West et al., 2002). A ranking system proposed by Lieveense et al. (2002) and utilised in subsequent systematic reviews (van Dijk et al., 2006; Tanamas et al., 2009) was modified for this purpose. Strong evidence was defined as generally consistent findings in multiple (≥ 2) high quality cohort studies; limited evidence required generally consistent findings in a single high quality cohort study; conflicting evidence resulted from conflicting findings in high quality studies (i.e. where $< 75\%$ of studies reported

consistent findings). Heterogeneity of included studies was examined prior to deciding on the method for synthesis of evidence.

2.3. Results

2.3.1. Search results

Results of the search strategy and screening process are shown in Figure 1. Consensus was achieved in all cases and the third reviewer was not required. Thirty studies were eligible for full review, twenty-five from the initial search and five from the update (Dougados et al., 1992; Schouten et al., 1992; Spector et al., 1992; Dieppe et al., 1993; Spector et al., 1994; Ledingham et al., 1995; McAlindon et al., 1996a; McAlindon et al., 1996b; Dieppe et al., 1997; Shiozaki et al., 1999; Cooper et al., 2000; Sharma et al., 2001; Wolfe et al., 2002; Bruyere et al., 2003; Sharma et al., 2003a; Sharma et al., 2003b; Felson et al., 2004; Thorstensson et al., 2004; Mazzuca et al., 2005; Mazzuca et al., 2006; Amin et al., 2007; Reijman et al., 2007; Amin et al., 2008; Amin et al., 2009; Helio Le Graverand et al., 2009; Niu et al., 2009; Benichou et al., 2010; Golightly et al., 2010; Harvey et al., 2010; Sharma et al., 2010).

2.3.2. Study details

Details of included studies are available in Appendix E. Participants were recruited from community settings, rheumatology or orthopaedic clinics, as well as previous studies. Sample sizes ranged from fifty-four (Thorstensson et al., 2004) to 2964 (Harvey et al., 2010). Four studies recruited females only (Spector et al., 1994; Shiozaki et al., 1999; Mazzuca et al., 2005; Helio Le Graverand et al., 2009); one study males only (Amin et al., 2007); with remaining studies reporting between 41% (Felson et al., 2004) and 84% female participants (Mazzuca et al., 2006). Mean age ranged from 44.8 (Thorstensson et al., 2004) to 70.3 years (McAlindon et al., 1996a; McAlindon et al., 1996b). KL grade was the most frequently used set of criteria for classification of knee OA. The American College

of Rheumatology (ACR) criteria and JSN score were also used. Follow up times varied from one year (Dougados et al., 1992; Helio Le Graverand et al., 2009; Benichou et al., 2010) to fourteen years (Shiozaki et al., 1999).

2.3.3. Assessment of Bias

Results of the bias assessment are shown in Table 2.2. Three reviewers assessed 600 items, agreeing on 503 (percentage agreement = 85%; $\kappa = 0.59$). Disagreements were resolved by discussion. The majority of studies (26/30) scored $\geq 60\%$ (≥ 12) and could be considered of high quality. Although few studies provided sample size calculation (item R), sample size was considered adequate for analysis in twenty-eight of the thirty included studies, based on a ratio of ten subjects per prognostic variable analysed (Concato et al., 1993, Childs and Cleland, 2006). The majority of studies were longitudinal cohorts, except four studies that used subjects recruited either partly or wholly from RCTs (Spector et al. 1992; Bruyere et al., 2003; Mazzuca et al., 2005; Mazzuca et al., 2006). This resulted in a good overall score for item I. Overall, criteria relating to study participation (A – D) were not well met. Reporting of response rates was lacking in many studies or confused due to lack of clarity about numbers approached, numbers eligible, and final numbers who participated (item D). Several studies presented only p-values to describe an association between predictive factor(s) and outcome (item P). Not all studies presented adjusted results for potential confounders, the minimum requirements for which were age and sex (item S). It should be noted that studies that scored highest addressed all aspects of bias to some extent; low quality studies failed to adequately address all areas of bias.

Sources of funding for studies were noted but not included in the scoring tool; five studies did not report this (Spector et al., 1992; Shiozaki et al., 1999; Bruyere et al., 2003; Mazzuca et al., 2006; Reijman et al., 2007). Results from all studies are included in Appendix F.

Table 2.2 Assessment of bias scores for included papers in systematic review

Author/Year	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	Score
(Sharma et al., 2010)	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	19
(Benichou et al., 2010)	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	18
(Sharma et al., 2003a)	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	18
(Sharma et al., 2003b)	1	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	18
(Sharma et al., 2001)	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
(Golightly et al., 2010)	1	1	1	0	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	17
(Harvey et al., 2010)	1	0	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	17
(Amin et al., 2008)	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	1	1	1	16
(Felson et al., 2004)	0	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	16
(Helio Le Graverand et al., 2009)	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	16
(Mazzuca et al., 2006)	1	0	0	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	16
(McAlindon et al., 1996a)	0	1	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16
(McAlindon et al., 1996b)	0	1	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16
(Niu et al., 2009)	1	1	0	0	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	16
(Spector et al., 1994)	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	16
(Amin et al., 2009)	0	0	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	1	1	1	15
(Cooper et al., 2000)	1	1	0	0	1	0	1	1	1	1	1	0	1	1	1	1	0	1	1	1	15
(Ledingham et al., 1995)	0	1	1	0	1	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	15
(Reijman et al., 2007)	0	1	0	1	1	0	0	1	1	1	1	1	1	1	1	1	0	1	1	1	15

Author/Year	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	Score
(Schouten et al., 1992)	1	1	0	0	1	0	0	1	1	1	1	1	1	1	1	1	0	1	1	1	15
(Thorstensson et al., 2004)	0	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	0	0	1	1	15
(Wolfe et al., 2002)	1	1	0	0	1	0	0	1	1	1	1	1	1	1	1	1	1	1	0	1	15
(Amin et al., 2007)	0	0	1	0	1	1	1	1	1	1	1	1	1	0	1	0	0	1	1	1	14
(Mazzuca et al., 2005)	1	1	0	0	1	1	0	1	1	1	1	1	1	1	1	0	0	1	0	1	14
(Shiozaki et al., 1999)	1	1	0	0	1	0	0	1	1	1	1	1	1	1	0	1	0	1	0	1	13
(Dieppe et al., 1997)	0	0	0	1	1	1	1	1	1	1	1	0	1	1	1	0	0	1	0	1	13
(Bruyere et al., 2003)	0	0	0	0	1	0	1	1	0	1	1	1	1	1	0	0	1	1	0	0	10
(Dieppe et al., 1993)	1	1	0	0	1	0	1	1	0	1	1	1	1	1	0	0	0	0	0	0	10
(Dougados et al., 1992)	0	0	0	0	0	0	1	1	1	1	1	1	1	0	1	0	0	1	0	1	10
(Spector et al., 1992)	0	0	0	0	1	0	0	1	0	1	1	0	1	1	1	0	0	1	0	0	8
Score/Item	15	20	13	7	29	17	16	30	26	30	30	27	30	26	26	20	13	28	22	26	

2.3.4. Synthesis of evidence

Included studies exhibited marked heterogeneity in terms of study design; descriptions of patient characteristics; follow-up periods; outcome measures; statistical analysis; and reporting of results. Consequently, meta-analysis was not possible and the alternative method was used.

Ten studies were excluded from the synthesis of evidence: four were of low quality (Dougados et al., 1992; Spector et al., 1992; Dieppe et al., 1993; Bruyere et al., 2003); six did not perform appropriate analysis to identify prognostic factors, or did not adjust for confounding (Sharma et al., 2003b; Mazzuca et al., 2005; Amin et al., 2007; Amin et al., 2008; Amin et al., 2009; Helio Le Graverand et al., 2009). Results from the remaining twenty studies are presented in Tables 2.3 and 2.4. Table 2.3 includes characteristics typically obtained from a patient history or self-administered questionnaires. Table 2.4 includes clinical variables that can be observed or measured by a health practitioner. Level of significance was set at $p < 0.05$ or confidence intervals that did not cross 1.0. Significant predictors of progression are identified as well as significant predictors of no-progression (which could be considered protective factors). Variables with no significant association with progression are also presented. Strength of evidence, based on the adapted ranking system, is stated for each variable. Number of studies, combined sample size, and study results give an indication of the magnitude of effect size, and the quantity of information from which the strength of evidence is obtained.

Patient characteristics with strong evidence for being predictors of knee OA progression are: older age; presence of OA in multiple joints; and varus malalignment of the knee. Evidence combined from studies using different radiographic measures suggests that radiographic features at baseline are strongly associated with progression. There was strong evidence that physical

activity or moderate participation in sport is not significantly associated with progression. All other variables had either limited or conflicting level of evidence.

Table 2.3 Evidence from high quality* studies for predictors of progression of knee osteoarthritis: demographics and patient history (self-report measures)

Baseline Variable	Level of Evidence†	Number of studies	Combined Sample size	Association with progression (95% CI)	Type of analysis Multivariate/ Univariate	Author/Reference
Predictor of progression						
Age	Strong	6	1006	OR = 3.84 (1.10, 13.4) OR = 1.34 (1.15, 1.57) OR 1.07 (1.02, 1.13) ↑ age and ↓ function more likely to progress ‡ β =0.01±0.006mm/year OR 1.13 (0.87, 1.48)§	Univariate Multivariate Multivariate Multivariate Multivariate Univariate	(Schouten et al., 1992) (Sharma et al., 2003a) (Ledingham et al., 1995) (Dieppe et al., 1997) (Benichou et al., 2010) (Mazzuca et al., 2006)
Sex	Conflicting	3	475	OR 2.17 (1.13, 4.15) male/↓pain more likely to improve than female/ ↑pain‡ female/↑rest pain/↓joint pain more likely to progress‡ OR 0.5 (0.22, 1.11)§	Multivariate Multivariate Multivariate Univariate	(Ledingham et al., 1995) (Dieppe et al., 1997) (Dieppe et al., 1997) (Schouten et al., 1992)
Pain	Conflicting	3	735	male/↓pain more likely to improve than female/ ↑pain‡	Multivariate	(Dieppe et al., 1997)

Baseline Variable	Level of Evidencet	Number of studies	Combined Sample size	Association with progression (95% CI)	Type of analysis Multivariate/ Univariate	Author/Reference
				OR 1.12 (0.90, 1.40)§ female/↑rest pain/↓joint pain more likely to progress‡§ OR 0.8 (0.4, 1.7)§ to OR 2.4 (0.7, 8.0)§	Multivariate Multivariate Univariate	(Sharma et al., 2003a) (Dieppe et al., 1997) (Cooper et al., 2000)
Function	Conflicting	3	487	↑ age and ↓ function more likely to progress ‡ OR 1.16 (0.92, 1.47)§ OR 0.98 (0.32, 3.02)‡§ to OR 1.56 (0.53, 4.60)‡§	Multivariate Univariate Univariate	(Dieppe et al., 1997) (Mazzuca et al., 2006) (Thorstensson et al., 2004)
Joint Stiffness	Limited	1	288	OR 1.39 (1.09, 1.77)	Univariate	(Mazzuca et al., 2006)
Disease Severity	Limited	1	1232	HR 1.01 (1.00, 1.02)	Multivariate	(Wolfe et al., 2002)
Symptom Duration	Limited	1	1232	HR 1.3 (1.00, 1.05)	Multivariate	(Wolfe et al., 2002)
Vitamin D Intake	Limited	1	126	OR 2.99 (1.06, 8.49) to OR 4.05 (1.4, 11.6)	Multivariate	(McAlindon et al., 1996a)
Predictor for not progressing (protective)						
Antioxidant intake:						

Baseline Variable	Level of Evidencet†	Number of studies	Combined Sample size	Association with progression (95% CI)	Type of analysis Multivariate/ Univariate	Author/Reference
↑Vitamin C	Limited	1	187	OR 0.06 (0.01, 0.67) to OR 0.31 (0.12, 0.79)	Multivariate	(McAlindon et al., 1996b)
↑ Vitamin E	Conflicting	1	187	OR 0.07 (0.01, 0.61) to OR 0.48 (0.40, 2.90)§	Multivariate	(McAlindon et al., 1996b)
Psychosocial factors:						
Mental health score	Limited	1	236	OR 0.58 (0.39, 0.86)	Multivariate	(Sharma et al., 2003a)
Social support	Limited	1	236	OR 0.85 (0.73, 0.98)	Multivariate	(Sharma et al., 2003a)
Self-efficacy	Conflicting	1	236	OR 0.80 (0.65, 0.98) to OR 0.86 (0.68, 1.09)§	Multivariate	(Sharma et al., 2003a)
No significant association with progression						
Physical activity/ sports participation	Strong	3	732	OR 0.84 (0.69, 1.02) to OR 0.86 (0.71, 1.05) OR 0.7 (0.4, 1.6) to OR 0.9 (0.3, 2.5) OR 0.53 (0.17, 1.68)§	Multivariate Univariate Univariate	(Sharma et al., 2003a) (Cooper et al., 2000) (Schouten et al., 1992)
Antioxidant intake: Beta-carotene	Limited	1	187	OR 0.26 (0.04, 1.57) to OR 1.75 (0.74, 4.11)§	Multivariate	(McAlindon et al., 1996b)

Baseline Variable	Level of Evidencet	Number of studies	Combined Sample size	Association with progression (95% CI)	Type of analysis Multivariate/ Univariate	Author/Reference
Psychosocial factors: Role functioning	Limited	1	236	OR 0.99 (0.75, 1.32)§	Multivariate	(Sharma et al., 2003a)
Previous knee injury	Limited	1	354	OR 1.2 (0.5, 3.0)§ to OR 1.1 (0.3, 4.4)§	Univariate	(Cooper et al., 2000)
Smoking	Limited	1	142	OR = 1.07 (0.38, 3.04)§ to OR = 0.96 (0.34, 2.75)§	Univariate	(Schouten et al., 1992)
Results ordered by number of studies, effect size and significance * Bias score ≥ 60% (≥12); † strong = consistent findings from ≥ 2 high quality studies, conflicting = < 75% high quality studies report consistent findings, limited = consistent findings from a single high quality study (Lievence et al., 2002); 95% CI = 95% confidence interval; OR = odds ratio; ‡ full data not given § non-significant; HR = hazard ratio						

Table 2.4 Evidence from high quality* studies for predictors of progression of knee osteoarthritis: physical examination and clinical tests

Baseline Variable	Level of Evidence†	Number of Studies	Combined Sample size	Association with progression (95% CI)	Type of analysis Multivariate/ Univariate	Author/ reference
Predictor of progression						
BMI - kg/m² > 27.3 25.97-27.73 > 27.5 >25.4 > 35 26.7-34.7 25.0-26.6	Conflicting	11	10, 084	OR 11.1 (3.28, 37.3) OR 5.28 (1.54, 18.1)	Univariate	(Schouten et al., 1992)
				OR 2.1 (1.2, 3.7) to OR 3.2 (1.1, 9.7)	Univariate	(Reijman et al., 2007)
				OR 2.6 (1.0, 6.8)	Univariate	(Cooper et al., 2000)
				RR 1.8 (1.0, 3.2)	Multivariate	(Niu et al., 2009)
				RR 1.51 (1.22, 1.87)‡ RR 1.38 (1.10, 1.73)‡	Univariate	(Shiozaki et al., 1999)
				OR 1.06 (1.00, 1.12) to OR 1.07 (1.02, 1.14)	Multivariate	(Ledingham et al., 1995)
				HR 1.03 (1.00, 1.06)	Multivariate	(Wolfe et al., 2002)

Baseline Variable	Level of Evidence†	Number of Studies	Combined Sample size	Association with progression (95% CI)	Type of analysis Multivariate/ Univariate	Author/ reference
25-34.9				$\beta = -0.03 \pm 0.01$ mm per kg/m ²	Multivariate	(Benichou et al., 2010)
				RR 4.69 (0.63, 34.75)§	Univariate	(Spector et al., 1994)
				RR 1.2 (0.7, 2.1) to RR 1.2 (0.7, 2.2)§	Multivariate	(Niu et al., 2009)
				OR 1.14 (0.89, 1.46)§	Multivariate	(Sharma et al., 2003a)
				OR 0.96 (0.72, 1.26) to OR 1.00 (0.83, 1.20)§	Multivariate	(Felson et al., 2004)
BMI: Stratified by alignment						
Moderate	Conflicting	1	227	OR 1.39 (1.07, 1.80) to OR 1.14 (0.92, 1.40)§	Multivariate	(Felson et al., 2004)
Severe	Limited	1	227	OR 0.92 (0.71, 1.18) to OR 0.96 (0.68, 1.36)§	Multivariate	(Felson et al., 2004)
Varus	Limited	1	2623	RR 0.9 (0.7, 1.1)§	Multivariate	(Niu et al., 2009)
Valgus	Limited	1	2623	RR 0.8 (0.6, 1.4) to RR 1.4 (0.9, 2.1)§	Multivariate	(Niu et al., 2009)

Baseline Variable	Level of Evidence†	Number of Studies	Combined Sample size	Association with progression (95% CI)	Type of analysis Multivariate/ Univariate	Author/ reference
Combined radiographic features: ↑Osteophyte score JSW JSN Severity: KL grade Chondrocalcinosis	Strong	4	2240			
		1	288	OR 0.47 (0.33, 0.66)	Univariate	(Mazzuca et al., 2006)
		1	288	OR 0.63 (0.47, 0.86)	Multivariate	(Mazzuca et al., 2006)
		1	1232	HR 2.53 (1.94, 3.31)	Multivariate	(Wolfe et al., 2002)
		1	188	OR 1.29 (1.07, 1.57) to OR 1.72 (1.36, 2.19)	Multivariate	(Ledingham et al., 1995)
1	532	OR 2.01 (0.55, 7.42)§	Univariate	(Schouten et al., 1992)		
Heberden's nodes/ Bony swelling/ Bouchard's nodes	Conflicting	3	684	OR 5.97 (1.54, 23.1) OR 1.80 (1.02, 3.17) OR 0.7 (0.4, 1.6) to OR 2.0 (0.7, 5.7)§	Univariate Multivariate Univariate	(Schouten et al., 1992) (Ledingham et al., 1995) (Cooper et al., 2000)
Multiple joint OA	Strong	2	720	OR 3.28 (1.30, 8.27) OR 2.39 (1.16, 4.93) to OR 2.42 (1.02, 5.77)	Univariate Multivariate	(Schouten et al., 1992) (Ledingham et al., 1995)
Alignment: Varus/medial JSN (or ↑KL grade)	Strong	2	1180	OR 2.98 (1.51, 5.89) to OR 4.09 (2.20, 7.62) OR 3.59 (2.62, 4.92)	Univariate Multivariate	(Sharma et al., 2001) (Sharma et al., 2010)
Valgus/lateral JSN	Conflicting	2	1180	OR 2.51 (0.91, 6.89)§ to	Univariate	(Sharma et al., 2001)

Baseline Variable	Level of Evidence†	Number of Studies	Combined Sample size	Association with progression (95% CI)	Type of analysis Multivariate/ Univariate	Author/ reference
(or ↑KL grade)				OR 4.89 (2.13, 11.20) OR 4.85 (3.17, 7.42)	Multivariate	(Sharma et al., 2010)
Bilateral	Conflicting	1	230	OR 2.23 (0.05, 4.41)§ to OR 3.22 (1.28, 8.12)	Univariate	(Sharma et al., 2001)
Unilateral	Limited	1	230	OR 0.17 (-1.66, 2.01) to OR 2.33 (0.97, 5.62)§	Univariate	(Sharma et al., 2001)
Leg length inequality	Conflicting	2	4547	HR 1.13 (0.53, 2.39)§ to HR 1.83 (1.10, 3.05) OR 1.0 (0.4, 2.5)§ to OR 1.3 (1.0, 1.7)	Multivariate Multivariate	(Golightly et al., 2010) (Harvey et al., 2010)
Ipsilateral PF OA	Limited	1	288	OR 2.31 (1.37 – 3.88) to OR 3.36 (1.83-6.18)	Univariate Multivariate	(Mazzuca et al., 2006)
Serum Vitamin D	Limited	1	126	OR 2.83 (1.02-7.85) to OR 2.89 (1.01-8.25)	Multivariate	(McAlindon et al., 1996a)
Warmth	Limited	1	188	OR 2.14 (1.30 – 3.52)	Multivariate	(Ledingham et al., 1995)
CPPD	Limited	1	188	OR 1.85 (1.04-3.29) to OR 2.41 (1.33-4.39)	Multivariate	(Ledingham et al., 1995)
↑volume synovial fluid	Limited	1	188	OR 1.02 (1.00-1.05) to OR 1.03 (1.01-1.05)	Multivariate	(Ledingham et al., 1995)

Baseline Variable	Level of Evidence†	Number of Studies	Combined Sample size	Association with progression (95% CI)	Type of analysis Multivariate/ Univariate	Author/ reference
Predictor for not progressing (protective)						
Alignment						
Varus/lateral JSN	Limited	1	950	OR 0.12 (0.07, 0.21)	Multivariate	(Sharma et al., 2010)
Valgus/medial JSN	Limited	1	950	OR 0.34 (0.21, 0.55)	Multivariate	(Sharma et al., 2010)
No significant association with progression						
Contralateral OA	Limited	1	288	OR 1.53 (0.82 – 2.85)§	Univariate	(Mazzuca et al., 2006)
Localised OA	Limited	1	532	OR 1.17 (0.51 – 2.72)§	Univariate	(Schouten et al., 1992)
Uric acid concentration	Limited	1	532	OR 1.05 (0.36-3.00) to OR 1.36 (0.46-4.02)§	Univariate	(Schouten et al., 1992)
Muscle strength	Limited	1	263	Quadriceps OR 0.88 (0.70-1.11)§ Hamstrings OR 0.86 (0.60-1.23)§	Multivariate Multivariate	(Sharma et al., 2003a) (Sharma et al., 2003a)
Results ordered by number of studies, effect size and significance						
* Bias score ≥ 60% (≥12); † strong = consistent findings from ≥ 2 high quality studies, conflicting = < 75% high quality studies report consistent findings, limited = consistent findings from a single high quality study (Lievence et al., 2002); 95% CI = 95% confidence interval; BMI = body mass index; OR = odds ratio; HR = hazard ratio; RR = relative risk ratio; ‡ full data not given; § non-significant results; JSW = joint space width; JSN = joint space narrowing; KL grade = Kellgren Lawrence grade; OA = osteoarthritis; PF = patellofemoral; CPPD = calcium pyrophosphate crystals						

2.3.5. Sensitivity analysis

Variation in length of follow-up in included studies was noted. As knee OA has been recognised as a slowly progressing disease (Felson, 1995; van Dijk et al., 2006), shorter follow-up times for studies may not allow sufficient time for progression to occur, resulting in smaller numbers for analysis of prognostic factors, a problem exacerbated in small cohort studies. Therefore a sensitivity analysis was performed synthesizing evidence from studies with a follow-up period of three-years or greater, a cut-off used in a previous review (van Dijk et al., 2006). Eleven studies met this criterion (Schouten et al., 1992; McAlindon et al., 1996a; McAlindon et al., 1996b; Dieppe et al., 1997; Shiozaki et al., 1999; Cooper et al., 2000; Wolfe et al., 2002; Sharma et al., 2003a; Thorstensson et al., 2004; Reijman et al., 2007; Golightly et al., 2010). BMI became a strong predictor of progression; evidence became limited for multiple joint OA; and conflicting for radiographic baseline features. Level of evidence for a number of baseline variables remained the same, while for others (including alignment) no evidence is available for the longer time period.

A sensitivity analysis was also performed based on quality. Findings of the review did not change if the quality threshold was reduced (< 60%), as no lower quality studies used appropriate analysis to identify prognostic variables. However, when the threshold for high quality studies was raised to $\geq 70\%$ (i.e. ≥ 15 with bias assessment tool for the current study), four articles were reclassified as low quality (Dieppe et al., 1997; Shiozaki et al., 1999; Mazzuca et al., 2005; Amin et al., 2007). Only two of these were previously included in the synthesis of evidence (Dieppe et al., 1997; Shiozaki et al., 1999) as the other two had been excluded due to inadequate analysis (Mazzuca et al., 2005; Amin et al., 2007). At the higher threshold of quality, there is strong evidence that baseline pain and function are not significantly associated with progression. Other findings are unaltered.

Results can also be synthesized according to type of outcome measure used. Age remains a strong predictor of progression when investigated using function as an outcome measure (Ledingham et al., 1995; Dieppe et al., 1997; Sharma et al., 2003a), while evidence becomes conflicting if radiographic outcome is used (Schouten et al., 1992; Mazzuca et al., 2006; Benichou et al., 2010). Other findings remain unchanged, excepting baseline function which changes to strong evidence of no association with progression when radiographic outcome is used (Thorstensson et al., 2004; Mazzuca et al., 2006).

2.4. Discussion

Baseline patient characteristics with strong evidence for being predictors of knee OA progression are: age, presence of OA in multiple joints (clinical observation), varus deformity of the knee (radiographic), and radiographic features (JSN/JSW, chondrocalcinosis, severity of OA as measured by KL grade, osteophyte score). Additionally, findings from sensitivity analyses suggest higher BMI is a strong predictor of progression over a longer time-period (> three years); and pain and function have strong evidence of no association with progression (quality score ≥ 15). These variables are easily assessed in clinical practice and may assist health care practitioners to provide their patients with appropriate advice and interventions for managing their disease. Patients can also be reassured that moderate participation in physical activity (patient report) is unlikely to have any effect on progression. Clinicians can have confidence in these findings as the information is drawn from high quality studies and exhibits consistency between studies. Furthermore, study samples were drawn from a variety of settings: orthopaedic and rheumatology clinics, as well as community- based cohorts, making findings of the review generalisable. However, quantity of information is somewhat limited for varus alignment, multiple joint OA, radiographic features, physical activity, knee pain, and function, due to low study numbers for evidence synthesis. Notably effect sizes for some variables are small or inconsistent. These

factors should be considered before applying review findings to clinical practice (West et al., 2002).

2.4.1. Comparison with previous research

Table 2.5 compares findings of the current and previous reviews for variables with strong evidence only. Similar findings from multiple reviews increase confidence in the conclusions. For example, strong evidence from two reviews, plus limited evidence from another review, identifies varus alignment as a predictor of progression. This also highlights the importance of biomechanical influences on radiographic progression of knee OA.

There is strong evidence from two reviews that generalised or multiple joint OA (assessed by clinical observation) is a predictor of progression. While the precise reasons for this association are not entirely clear, it is possible that generalised OA may reflect an underlying systemic or genetic influence on cartilage that contributes to increased likelihood of disease progression (Schouten et al., 1992).

Similarly, two reviews found strong evidence that participation in physical activity is not significantly associated with progression, with limited evidence from another review. Physical activity included aerobic exercise, jogging, or being a member of a sports club (Schouten et al., 1992; Cooper et al., 2000; Sharma et al., 2003a). This lack of association with disease progression is helpful given the many recognized health benefits associated with regular exercise.

Table 2.5 Comparison of findings from systematic reviews that identify variables with strong evidence for progression of knee osteoarthritis

Baseline Variables	Current Study	Belo et al. (2007)	van Dijk et al. (2006)	Tanamas et al. (2009)
Predictors				
Knee malalignment	Strong (varus alignment with medial JSN)*	Limited	Limited †	Strong
Multiple joint/ generalised OA	Strong	Strong	-	-
Age	Strong	Conflicting	Limited	
Radiographic features	Strong	Strong †	-	-
Hyaluronic acid	-	Strong	-	-
Not associated with progression				
Participation in physical activity	Strong	Strong	Limited	-
Muscle strength	Limited	Strong	Limited (progression less likely)	-
Previous knee injury	Limited	Strong	-	-
Unclear association with progression				
Sex	Conflicting	Strong †	Limited	-
Knee pain	Conflicting	Strong †	Limited (predictor)	-
JSN = joint space narrowing; *other results limited/conflicting; OA = osteoarthritis; † not associated				

There are inconsistent findings for the remaining variables in Table 2.5. For example in the current review, combined radiographic features have strong evidence of being predictive of progression. In contrast, Belo et al. (2007) report strong evidence that radiologic severity is *not* associated with progression, having arrived at a different interpretation from similar studies.

Dissimilar review objectives may have contributed to disparity in findings. Belo et al. (2007) included articles with progression defined by radiographic measures only. Unlike the current review they did not require inclusion of clinical features at baseline. Tanamas et al. (2009) conducted a review focussed solely on the role of alignment as a predictor; Van Dijk et al. (2006) concentrated on studies with changes in functional status or pain. Additionally the pool of articles in each review was affected by decisions such as date of search, search strategies, and inclusion/exclusion criteria. Inclusion of different source studies contributes to diversity of conclusions in the reviews. For example age was found to be a strong predictor in the current review and conflicting in another. However, there was only one study that contributed results on age to both reviews (Schouten et al., 1992), with other results taken from different studies.

2.4.2. Study limitations

As indicated above, number of studies will also affect the strength of evidence and may have contributed to the disparity in findings between reviews. Relatively few studies were included in the current review for variables such as function and pain (three studies each), compared to the number that examined BMI (eleven studies) or age (six studies). This makes synthesis of evidence susceptible to change with the addition or deletion of articles from the analysis. This point is reinforced by results of a sensitivity analysis which raised the threshold for high quality studies to $\geq 70\%$. The exclusion of one study (Dieppe et al., 1997) changed the evidence from conflicting to strong, that pain and function were not significantly associated with progression. Tables 2.3 and 2.4 highlight the low number of studies from which the level of evidence is calculated.

There was a wide range of outcome measures used in included studies. Results of a sensitivity analysis revealed that age remained a strong predictor when assessed with change in function (Ledingham et al., 1995; Dieppe et al.,

1997; Sharma et al., 2003a); however, evidence became conflicting when using radiographic outcome measures (Schouten et al., 1992; Mazzuca et al., 2006; Benichou et al., 2010), exemplifying how choice of outcome measure for knee OA progression can influence results. However, there are no universally accepted outcome measures and most have recognised limitations. For radiographs these include the slow rate of progression and the fact that early disease may not be manifested in radiographic features (Kraus, 2006; Hunter et al., 2009a). Outcome measures describing change in pain or function lack responsiveness, with limited evidence that they deteriorate over a three-year period (van Dijk et al., 2006). Levels of pain or function may reflect short-term fluctuations in disease activity rather than progression. Alternatives such as magnetic resonance imaging (MRI) changes, or biomarkers were not utilised in this review due to uncertainty about their responsiveness in detecting progression, lack of standardised or universally accepted scoring methods, and their lack of clear relationship to clinical symptoms (Kraus, 2006; Hunter et al., 2009a). In addition they may lack clinical utility due to cost and technical requirements, being more suited to use in the research environment.

This review's focus was to identify predictive variables that could be easily assessed in routine clinical practice. This resulted in exclusion of some studies that used sophisticated equipment or analytical techniques, making it possible that other predictors of knee OA progression have been missed. Other studies were excluded as they described chronic knee pain, rather than diagnosed knee OA at baseline (Peters et al., 2005; Thomas et al., 2008). These samples probably represented individuals at an early stage of disease, and it is possible that predictors could be different in this patient population.

Other limitations of the current review relate to shortcomings of included studies. Prognostic studies have been criticised for poor quality and variable

methodology (Altman, 2001). The potential impact on the current findings was addressed with the assessment of bias tool. Results in Table 2.2 suggest conclusions from this review could be at risk of selection bias and attrition bias due to weaknesses in included studies.

2.4.3. Implications for future research

Improving selection and reporting of study participants, especially response rates at different stages, and reasons for dropouts, would address these biases and should be incorporated into future research. Two thirds of reviewed studies addressed confounding by age and sex, although there may be other potential confounders not investigated in these studies, so it is not possible to conclude definitively that confounding was absent (Simunovic et al., 2009). However, this review's requirement for adjusted or multivariate analyses should have minimised the effect of confounding.

The lack of standardised methods of assessment for baseline variables in addition to the variety of outcome measures already outlined, are two areas that should be addressed in future research. Consistent measures would make pooling of results and meta-analysis more likely and thereby strengthen the body of evidence and confidence in the findings. This approach may need to be tempered if new methods for defining progression are validated and adopted as best-practice (e.g. MRI changes or biomarkers) (Kraus, 2006; Hunter et al., 2009a).

Identification of predictors of long term progression is an area requiring further research. Existing, well-designed and adequately powered studies can indicate known variables worthy of further investigation. These variables should be included in future prognostic studies, with careful analysis of any added benefit of new predictors (Centre for Reviews and Dissemination, 2008).

2.4.4. Summary

This review has summarised current evidence for baseline characteristics that predict progression of knee OA. All of the variables identified with strong evidence as predictors (varus alignment; presence of OA in multiple joints; age; radiographic features; BMI) can be easily evaluated and utilised in clinical practice. Additionally knowledge that participation in physical activity is not associated with progression should be used to encourage patients to remain active. Patients can also be reassured that presence of baseline pain and limitation of function is not associated with progression. There are numerous other potential predictors where current evidence is limited or conflicting.

2.5. Implications for selection of baseline variables as potential predictors of treatment response for patients with knee OA

Variables identified in the review have a role in predicting progression of knee OA; in contrast the aim of this thesis is to identify predictors of successful response to treatment, which is arguably the opposite of progression. However, any variables with strong influence on outcome, especially if potentially modifiable by physiotherapy intervention, could possibly be identified as predictors of treatment response. The review has identified variables that are associated with outcome and this knowledge will be used to guide selection of variables for the baseline assessment in the main prognostic study. In addition to variables with strong evidence from a number of studies, variables with reported large ORs, RRs, or HRs will be considered for inclusion, even if overall evidence for prediction of progression was limited or conflicting; for example presence of ipsilateral patellar-femoral OA (Mazzuca et al., 2006) and psychosocial factors (Sharma et al., 2003a) (Tables 2.3 and 2.4).

Moreover, comparison of findings from the review and the main prognostic study should aid discrimination between predictors of natural progression and

variables that predict treatment response. A requirement of the review that predictors could be easily evaluated in clinical practice allows for their inclusion in the main prognostic study, which has a similar focus.

3. The concept of irritability and development of the i-score

3.1. Overview

This chapter will explain the concept of irritability, its importance, and how it is used in routine clinical practice. The potential for irritability to be a baseline predictor of outcome for patients with knee osteoarthritis (OA) will be discussed, and therefore its place in the current research. There is a lack of a standardised assessment and reporting of irritability for both clinical practice and research purposes. Methods for developing and testing a scale to meet this need are described. The resulting scale (the i-score) is presented and its psychometric properties discussed. Finally possible uses of the new scale are considered, as well as use of the i-score as a potential predictor of outcome for patients with knee OA following physiotherapy intervention.

3.2. Concept of irritability

Irritability is a component of symptom behaviour and describes the degree to which a musculoskeletal condition is exacerbated by movement. It is usually defined in terms of the duration or vigour of an activity or posture which produces the symptom, the severity of the symptom when present, and the ease with which the symptom can be alleviated when the activity or posture ceases (Maitland, 1991; Refshauge and Latimer, 1995). In the area of musculoskeletal physiotherapy and especially manual therapy it has been conceptualised and developed by Maitland (1986). It is an important factor for predicting the tolerance of the patient to evaluation and treatment procedures, although its reliability and validity has not been established (Cyriax, 1982; Maitland, 1986; Maitland, 1991; Refshauge and Latimer, 1995).

Irritability is a familiar concept to physiotherapists who consider it when assessing and treating patients (Hurley et al., 2002; Smart and Doody, 2007). It is

important to accurately assess and predict patient response to movement when planning a physical examination, to prevent unwarranted exacerbation of symptoms. Similarly, consideration of irritability should influence the vigour with which treatment techniques are applied: sufficient to promote healing while still preventing aggravation of symptoms or unwanted complications (Maitland, 1991; Hurley et al., 2002). Conversely if irritability is overestimated, assessment techniques may not stress structures sufficiently to identify the source of symptoms. This will cause difficulty with diagnosis and result in ineffective treatment, delaying recovery and adversely affecting patient outcomes (Maitland, 1991). In these contexts it is evident that irritability is used by physiotherapists as a predictor of patient response, at least in the short-term (Hurley et al., 2002; Smart and Doody, 2007). However, irritability as a predictor of response over longer periods, such as the one year follow-up in the main prognostic study, has not been investigated. Additionally, the role of irritability in the physiotherapy management of knee OA has not previously been considered. For the reasons outlined above it was hypothesised that irritability has the potential to influence treatment outcome and therefore warranted further investigation.

An issue that became clear early in the investigation was that assessment and reporting of irritability lacked clarity, with no standardised process or established guidelines available. Assessment methods have been variable, and considered irritability as either falling on a continuum ranging from 'irritable' to 'non-irritable', or categorizing it as 'low', 'moderate' or 'high', based on the patient response to questioning about site of pain, aggravating and easing factors, pain at night, functional limitations and intensity of pain (Cyriax, 1982; Maitland, 1986, 1991; Refshauge and Latimer, 1995). In routine clinical practice assessment of irritability is subject to variation, based on an individual clinician's interpretation of the relative importance of each of these items. A recent study investigated the inter-therapist reliability of physiotherapists' assessment of

irritability in a low back pain population (Barakatt et al., 2009a). This found only “fair to moderate” levels of agreement, which the authors suggested was due to lack of a standardized operational definition for judging irritability. Similar difficulties arose when trying to assess irritability for current research purposes.

3.3. Scale development and testing

A possible solution to this gap in current knowledge was to develop a new scale. The development of a scale is not without its difficulties and limitations. However, the advantages of having a standardised method for recording and reporting irritability include: facilitation of communication between physiotherapists by quantifying the degree or severity of irritability, and improvement of inter-rater reliability. Additionally, it will enhance teaching of students and novice physiotherapists about the concept, and highlight the usefulness of irritability assessment in clinical practice. Provision of a valid and reliable measurement tool is also important for research purposes.

Development of a scale should conform to a process outlined by Streiner and Norman (2003) involving: item selection, scaling of responses, conversion of items to a scale and finally testing the psychometric properties of the newly developed scale.

3.3.1. Item selection

An important step in the process of item selection or development is identification and evaluation of existing measures (DeWalt et al., 2007). Making use of expert judgements and clinical observation is also accepted practice and incorporates accumulated knowledge and expertise in the field (Streiner and Norman, 2003).

3.3.2. Scaling of responses

This includes decisions about the number of responses that can be made to each question or item in the scale. There is some debate about the ideal number of

responses, with fewer responses resulting in loss of information, and decreased reliability and accuracy (Streiner and Norman, 2003; DeWalt et al., 2007). A larger number of responses can increase the “cognitive burden” on respondents (DeWalt et al., 2007), in addition to becoming more time-consuming to complete. A general recommendation is to use between five and nine steps (Streiner and Norman, 2003). A noteworthy point is that measurement scales for individual items are frequently ordinal with no guarantee that divisions between categories are of equal value (Streiner and Norman, 2003). Numerical rating is the preferred method for assigning scores to responses (Streiner and Norman, 2003).

3.3.3. Conversion of items to a scale

The most straightforward method for converting item responses to a scale is to add the scores for each item together to get a total score. The benefit of weighting items to reflect their relative importance to the final score is debatable, with the increased complexity described as “rarely worth the effort” (Streiner and Norman, 2003).

3.3.4. Testing psychometric properties

Demonstration of validity is essential as it indicates that the scale is measuring what it is intending to measure. Streiner and Norman (2003, p.174) stated:

“Validating a scale is really a process whereby we determine the degree of confidence we can place on inferences we make about people based on their scores from that scale”

Different aspects of validity can be investigated.

Content validity

Content validity considers if there has been appropriate coverage of the area and also relevance of included items (Bland and Altman, 2002).

Criterion validity and correlation with existing measure of irritability

Criterion validity looks at how well results compare to an existing “gold standard” measure (Bland and Altman, 2002; Streiner and Norman, 2003). A strong correlation between a new scale and the “gold standard” measure indicates good criterion validity. However, absence of an existing standardised measure is frequently the justification for developing a new scale. In such cases comparisons may be made with whatever current measure is being used. Correlation between measures should still be detectable, although is likely to be lower.

Construct validity

Construct validity considers underlying theories or concepts and how well they are assessed by a scale or instrument (Bland and Altman, 2002; Streiner and Norman, 2003). Evaluating construct validity should involve testing relationships between the results of a new scale and other clinical variables measuring similar constructs (Bland and Altman, 2002). Correlation is used to evaluate this relationship.

Inter-rater reliability

A clinical measure must be reliable as well as accurate if it is to be useful for making clinical decisions. Bland and Altman (2002) suggest that if two measures are comparable in terms of validity, the more repeatable one should be selected for use.

Inter-rater reliability should be investigated to determine if a scale can be used by different testers and still produce a consistent result. High inter-rater reliability increases confidence in the accuracy of a scale.

Internal consistency

Internal consistency is a frequently reported property and reflects homogeneity of a scale (Streiner and Norman, 2003). *Individual items* in the scale should measure similar underlying symptoms or behaviour. They should relate

to each other, without too much repetition or overlapping, which would make some items redundant (Bland and Altman, 2002). *The total score* should also reflect the underlying concept and correlate with individual items (Streiner and Norman, 2003). Cronbach's alpha (α) is a useful summary measure of internal consistency and is used frequently in scale development (Bland and Altman, 2002; Streiner and Norman, 2003). Ideally α should be greater than 0.7 indicating good internal consistency and less than 0.9 to avoid redundancy of items (Nunnally, 1978).

Therefore the aim of this investigation was to develop a scale for the measurement of irritability that considered these aforementioned factors. This could be used subsequently to determine if irritability was a predictor of outcome at one year for patients with knee OA following physiotherapy intervention.

3.4. Methods

This study was nested within a randomised controlled trial (RCT) investigating the efficacy of physiotherapy interventions for lower limb osteoarthritis OA (Abbott et al., 2009).

3.4.1. Existing method for assessment of irritability

Five physiotherapists providing interventions in the RCT recorded irritability for each trial participant at the beginning of each treatment session. No instructions were given on the procedure for assessing irritability; rather it was left to the individual physiotherapists' clinical judgement and usual practice. Irritability was recorded as low, moderate or high.

3.4.2. Development of scale for assessment of irritability

Item selection

Four items were selected for inclusion in the scale based on the expert judgement of experienced physiotherapists and reflecting key components of irritability identified in the literature (Maitland, 1986; Maitland, 1991; Refshauge

and Latimer, 1995). These were aggravating activity, time to onset or worsening of pain, intensity of pain, and time for return to resting level of pain.

The scale was developed specifically for patients with hip and knee OA, therefore aggravating activities included those associated with lower limb OA, which are frequently assessed in quality of life questionnaires such as the Western Ontario and McMaster Universities Osteoarthritis index (WOMAC) function subscale (Bellamy et al., 1988). Hip OA was included as many aggravating activities are similar to those for knee OA and existing tools usually address both conditions (e.g. WOMAC). Included activities covered the range of performance likely to occur at different stages of disease, and accounted for factors such as raising or lowering of centre of gravity (COG), muscle action, joint loading, and range of motion. Also forces through the lower limb joints will be affected by the use of walking aids, carrying weights, and momentum or speed of performance.

Time to onset and time to relief of pain were recorded as categories, reflecting how patients typically report duration.

Pain intensity was recorded using the 11-point numeric pain rating scale (NPRS). This had the advantage of familiarity to participants in the clinical trial and has been shown to be reliable for use with older people with musculoskeletal pain (Jensen et al., 1999; Mawdsley et al., 2002; Abbott et al., 2009).

These four items were included in the baseline assessment for the RCT (Appendix H).

Scaling of responses

A decision was made to assign equal weight to each item. As the NPRS has a maximum score of ten, this meant that each of the other items also had a maximum score of ten.

Operational definitions for categories of aggravating activities were based on the criteria mentioned above. There were four time categories assigned, based on the question included in the RCT baseline assessment (Appendix H, p. 6). The

minimum response of both time to onset and time to easing of symptoms, therefore, scored two. Categories were ordered so that higher scores reflected increasing severity.

Conversion of items to scores – the i-score

Scores from each of the four items described were combined to give a single score for irritability. Minimum score was four and maximum score forty, with a higher score representing higher irritability. The author calculated the i-score from the baseline assessment of the RCT, for all knee participants and hip participants involved in psychometric testing.

3.4.3. Testing psychometric properties of the i-score

Content validity: evaluated by nominal group technique

Items included in the i-score were compared to top-ranked items identified by physiotherapists as being important in the assessment of irritability, using nominal group technique (NGT). NGT is a qualitative research technique that is quick and easy to perform, provides immediate feedback and uses consensus to reach its conclusions (Potter et al., 2004).

An NGT session was conducted by the author and an assistant. No detail of method or content was given in advance. On arrival physiotherapists were welcomed and thanked for their participation. It was emphasised that all contributions would be valued. The NGT method was described to them in detail and they were provided with an outline of the session (Figure 3.1). A carefully worded question was provided with little additional information.

The question was,

“With regard to patients with osteoarthritis of the hip or knee in the MOA trial, what factors do you think should be used to determine joint irritability?”

There were no definitions of terms so as not to bias responses. The NGT session was conducted as described in Figure 3.1.

<p>Step1. Opening statement Welcome and introductions. To set scene, acknowledge everyone’s attendance and everyone’s contribution. Overall aim: to produce consensus about a couple of issues key to thesis study. Describe procedure</p> <p style="text-align: right;">(10 mins)</p>
<p>Step 2. Silent Generation of Ideas. Consider question and record all thoughts/ideas/responses. Can be free thinking and wide ranging, no limits (as long as relates to question) No discussion, work silently and independently.</p> <p style="text-align: right;">(10 mins)</p>
<p>Step 3. Round Robin Recording of Ideas. Each physiotherapist in turn will be asked to contribute 1 idea/point. Keep to one thought at a time, a short phrase or sentence, no elaboration or justification. No comments from the floor! The next physiotherapist will then be invited to contribute a comment/idea. Goes in turns all around the room until all ideas are exhausted. A person may pass their turn if all their ideas have been expressed. They can re-enter if they get new ideas based on others contributions. The facilitator will record the ideas or comments on a whiteboard or flipchart, using the physiotherapists’ own words.</p> <p style="text-align: right;">(15 mins)</p>
<p>Step 4. Clarification of Ideas. Time for open discussion, clarification of points. Physiotherapists can agree if some items are duplicating an idea and group them together. No items can be deleted and the ideas should not be grouped into general categories where the detail may be lost. New items can be added as a result of discussion. The facilitator should encourage open and non-judgmental discussion; ensure all physiotherapists have an equal opportunity to contribute and that no particular individual is allowed to dominate discussion.</p> <p style="text-align: right;">(2 mins per item, 20 mins approx)</p>
<p>Step 5. Voting and Ranking. From the list each physiotherapist selects their 5 most important items. Then they rank them with the most important ranked highest.</p> <p style="text-align: right;">(5 mins)</p> <p>Voting forms collected and scores tallied. Results are given immediately on the scores for each item/idea. Feedback is invited from physiotherapists as to level of agreement. If necessary a further round of scoring can be done from the initial results.</p> <p style="text-align: right;">(10 mins)</p>

Figure 3.1 Procedure for nominal group technique

After the initial ranking of ideas by each physiotherapist, the assistant gathered the ranking sheets and collated the overall scores for each item. The group was invited to review the scores and ranking. Further clarification was required due to the overlap of ideas in different items and some confusion over terminology. Then another round of ranking took place. Results were collated and displayed. Further discussion led to consensus for the top five items physiotherapists considered most important to determine irritability. These items were correlated with the items included in the i-score to determine content validity.

Criterion validity: correlation with existing method

The absence of an existing standardised measure of irritability was the reason for developing the i-score. However, for the RCT physiotherapists were asked to assess and record irritability for all participants according to their usual practice (as outlined in section 3.4.1). The rating of irritability from the first treatment session was correlated with the i-score to determine criterion validity.

Construct validity: correlation with pain measures

To test construct validity of the i-score, results were correlated with pain reports for two physical performance measures (PPM), and pain fluctuation over a twenty-four hour period. Timed up and go (TUG) and thirty second sit to stand were the two PPMs (Podsiadlo and Richardson, 1991; Guralnik et al., 1994). In addition to recording time or repetitions, participants were asked to rate their pain using an eleven-point NPRS following completion of the task; thereby associating activity and pain response, one of the underlying constructs of irritability. A third measure involved calculating the difference in the best and worst pain in a twenty-four hour period, reflecting variability in pain behaviour, another construct of irritability. As before, information was extracted from baseline assessment in the RCT for both pain measures and the i-score.

Inter-rater reliability

As a potential predictor of treatment response, irritability needs to have demonstrated inter-rater reliability for widespread use by many investigators. Participants in the RCT consented to an additional investigator repeating four questions relating to irritability from the baseline assessment. This additional testing was performed immediately on completion of the original assessment. Duration and variety of procedures involved in the baseline assessment served to lengthen the time between questioning of irritability by each rater, and hopefully decrease memory or recall of responses by the participant. The second rater was blind to the responses gained by the first rater.

The two raters were physiotherapists involved in the Management of Osteoarthritis (MOA) trial. No training was required as the questions involved are self-explanatory.

Internal consistency

This was investigated to identify the degree to which the four items and the total of the i-score related to each other and the underlying construct of irritability.

3.4.4. Recruitment

All participants recruited for the RCT completed a baseline assessment, including the four questions comprising the i-score. Data were available for both hip and knee OA participants. The author calculated the i-score for all knee participants and for hip participants as required for psychometric testing.

All five physiotherapists delivering treatment in the RCT were invited to participate in the NGT study to investigate content validity of the i-score. Four physiotherapists were able to participate, one declined due to previous commitments. Participating physiotherapists ranged in age from twenty-six to fifty years; post-registration clinical experience ranged from six years to twenty years. Qualifications and training level varied, but all were New Zealand

registered physiotherapists working in musculoskeletal clinical practice. They were not involved in the construction of the i-score and were blinded to the results generated from the i-score.

All participants randomised to treatment groups were assessed for irritability by the same five physiotherapists at their first treatment session, using existing clinical methods. A subgroup of participants (n = 83) with hip or knee OA was constructed to test criterion validity. Another subgroup of participants (n = 109) was constructed to test for construct validity. Testing was completed during the conduct of the RCT accounting for different sample sizes, as more participants were recruited.

A sample of convenience (n = 28) to test for inter-rater reliability was constructed from consecutive participants, with either hip or knee OA, attending for baseline assessment in the clinical trial.

3.4.5. Data analysis

All statistical analyses were performed using STATA 10.1 (StataCorp, 4905 Lakeway Drive, College Station, Texas 77845, USA).

Analyses of Pearson correlation coefficients, and associated levels of significance, were used to determine criterion and construct validity.

Inter-rater reliability was determined using intraclass correlation coefficients (ICC), with corresponding standard errors of measurement (SEM) (Portney and Watkins, 2000). As the participants were assessed by two raters representative of the usual population of physiotherapists, and the results were single scores (not mean scores), ICC model 2,1 was used (Portney and Watkins, 2000). The SEM quantifies the measurement error in the same units as the original measurement, thus providing a clinically meaningful value to be interpreted by the clinician (Stratford and Goldsmith, 1997). Ninety-five percent confidence intervals (95% CIs) were calculated.

Internal consistency of the i-score was evaluated using Cronbach's α .

3.5. Results

Scale development resulted in the i-score, a measure to assess and report irritability for individuals with knee or hip OA, which is presented in Table 3.1.

Minimum possible score for irritability was four and maximum was forty, with higher values representing higher irritability.

Table 3.1 The i-score: a scale for assessment and reporting of irritability for patients with knee or hip osteoarthritis

Activity	Score A	Time to pain onset	Score B	NPRS	Score C	Time to pain relief	Score D
No limit	0	unlimited	2	0	0	Immediate - < 1min	2
				1	1		
FWB + weights/ vigorous activity	2	> 60 min	4	2	2	1 min - < 10 min	4
				3	3		
FWB repeated/ prolonged activity change in COG	4	10 – 60 min	6	4	4	10 – 60 min	6
				5	5		
FWB > PWB single movement change in COG with help	6	1 - < 10 min	8	6	6	> 60 min	8
				7	7		
PWB > FWB no change in COG short duration activity	8	Immediate - < 1 min	10	8	8	Doesn't settle	10
				9	9		
NWB limited movement or activity	10			10	10		
Subtotals /10							
Total Score (A + B + C + D) /40							
NPRS = numeric pain rating scale; FWB = full weight bearing; COG = centre of gravity; PWB = partial weight bearing; NWB = non weight bearing							

The results of applying the i-score to baseline data for 109 participants with knee OA are shown in Table 3.2 and Figure 3.2. There was incomplete data for four participants.

Table 3.2 Descriptive statistics for the i-score applied to all knee osteoarthritis participants

Participants	Mean	Standard Deviation	Minimum score	Maximum score
109	24.22	3.58	15	34

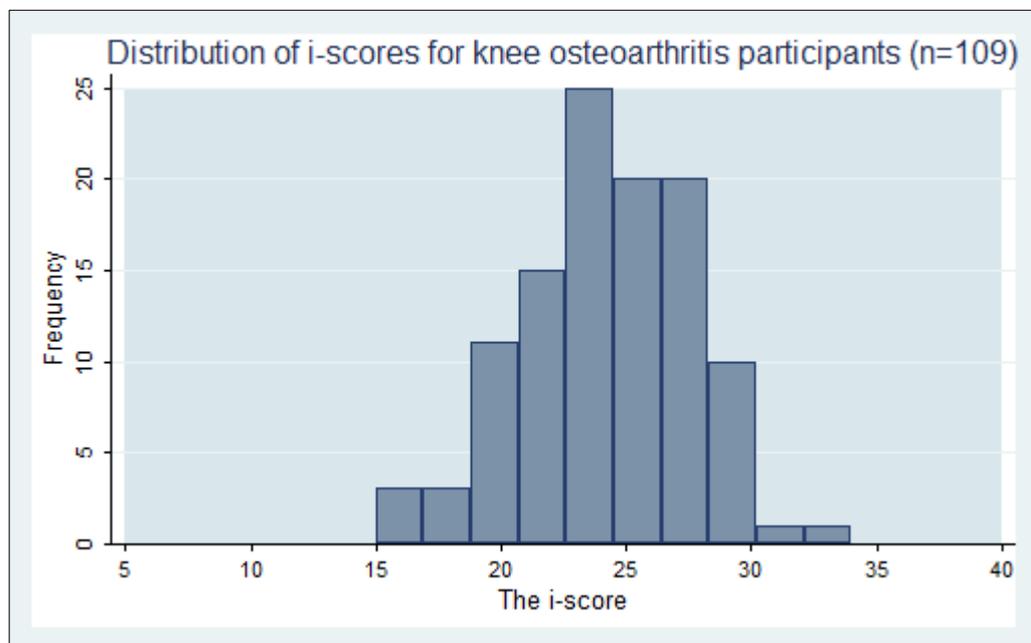


Figure 3.2 Histogram showing the distribution of the i-score in knee osteoarthritis participants

Operational definitions were provided to aid understanding and categorizing of aggravating activities (Table 3.3).

Table 3.3 Operational definitions: aggravating activities for knee and hip osteoarthritis

No limit	Free unconstrained movements, no limit to activity
FWB+	Slight limit to activity e.g. if weights added (up/down ladders, up/down activity with weight)
FWB	Activity limited when raising body weight, prolonged or repeated activity (up/down stairs or hills, walking prolonged duration/uneven ground, repeated squat)
FWB > PWB	Activity limited when raising body weight with assistance e.g. use of arm rests/handrail/furniture, single movements (rising from sitting, getting up from floor, stand and twist/turn, kneeling, single squat/bend to floor)
PWB > FWB	Weight bearing activity limited, no change in height of COG (standing, walking around house/short duration/short distance, walking with stick/frame)
NWB	Activity limited in NWB position (sitting/lying/changing position)
FWB = full weight bearing; PWB = partial weight bearing; COG = centre of gravity; NWB = non weight bearing	

Content validity using NGT

Two rounds of NGT were necessary to reach consensus about the most important factors for assessing irritability in patients with knee or hip OA. Ten factors were identified in the first round, decreasing to six in the second round. In the second round both pain intensity and duration of pain (“how long does it last?”) received nine points (see Table 3.4). A related item, level of symptoms, included pain, so following discussion it was agreed to rank pain intensity higher than pain duration. Final ranking was acceptable to all participants, therefore consensus was reached. Four of the top five ranked items were the same as items

in the i-score. The exception was impairment of function when pain is present (“what can you do when it is there?”). Time taken to complete the NGT session was ninety minutes. Detailed results are given in Table 3.4

Table 3.4 Results of Nominal Group Technique to decide important factors for assessing irritability for knee and hip osteoarthritis

	Factors affecting irritability	Score*	Participants reporting factor
Round 1			
	How easily is pain provoked? (activity)	14	4
	Pain intensity/pain	9	2
	What can you NOT do when pain is there?	9	3
	How long does pain last? (time)	7	4
	Rate of pain onset (time)	5	1
	Latency (delayed onset pain)	4	2
	Level of symptoms (pain and other)	4	1
	Functional activities that aggravate/ease	4	1
	Severity of symptoms compared to precipitating incident	3	1
	Response to treatment	1	1
Round 2			
	How easily is pain provoked? (activity)	20	4
	What can you NOT do when pain is there?	11	4
	Pain intensity/pain	9	4
	How long does pain last? (time)	9	4
	Rate of pain onset (time)	6	3
	Level of symptoms (pain and other)†	5	2
*Total of scores from all participants			
†Item excluded from top five ranked factors			

Criterion validity

Eighty-five of the total sample of participants with knee OA were randomised to one of the treatment arms of the RCT. Physiotherapists rating of

irritability was recorded for eighty-three of these at the first treatment session, with missing data for two participants. Physiotherapists rated thirty-nine participants as low irritability; forty-two as moderate irritability; and two participants with high irritability. For the same eighty-three participants descriptive statistics are shown in Table 3.5.

Table 3.5 Descriptive statistics for the i-score applied to participants in the treatment arms of a clinical trial

Participants	Mean	Standard Deviation	Minimum score	Maximum score
83	24.25	3.46	16	32

Analysis showed a moderate level of correlation between the i-score and physiotherapists' existing method for assessment of irritability ($r = 0.48$). This finding was significant ($p < 0.0001$).

Construct validity

These analyses were conducted on 109 participants with data available from the baseline assessment. Figure 3.3 shows the distribution of pain scores from two physical performance measures (TUG and Sit to Stand) and the change in pain over a 24-hour period (using NPRS). Descriptive statistics for the same 109 participants are shown in Table 3.2.

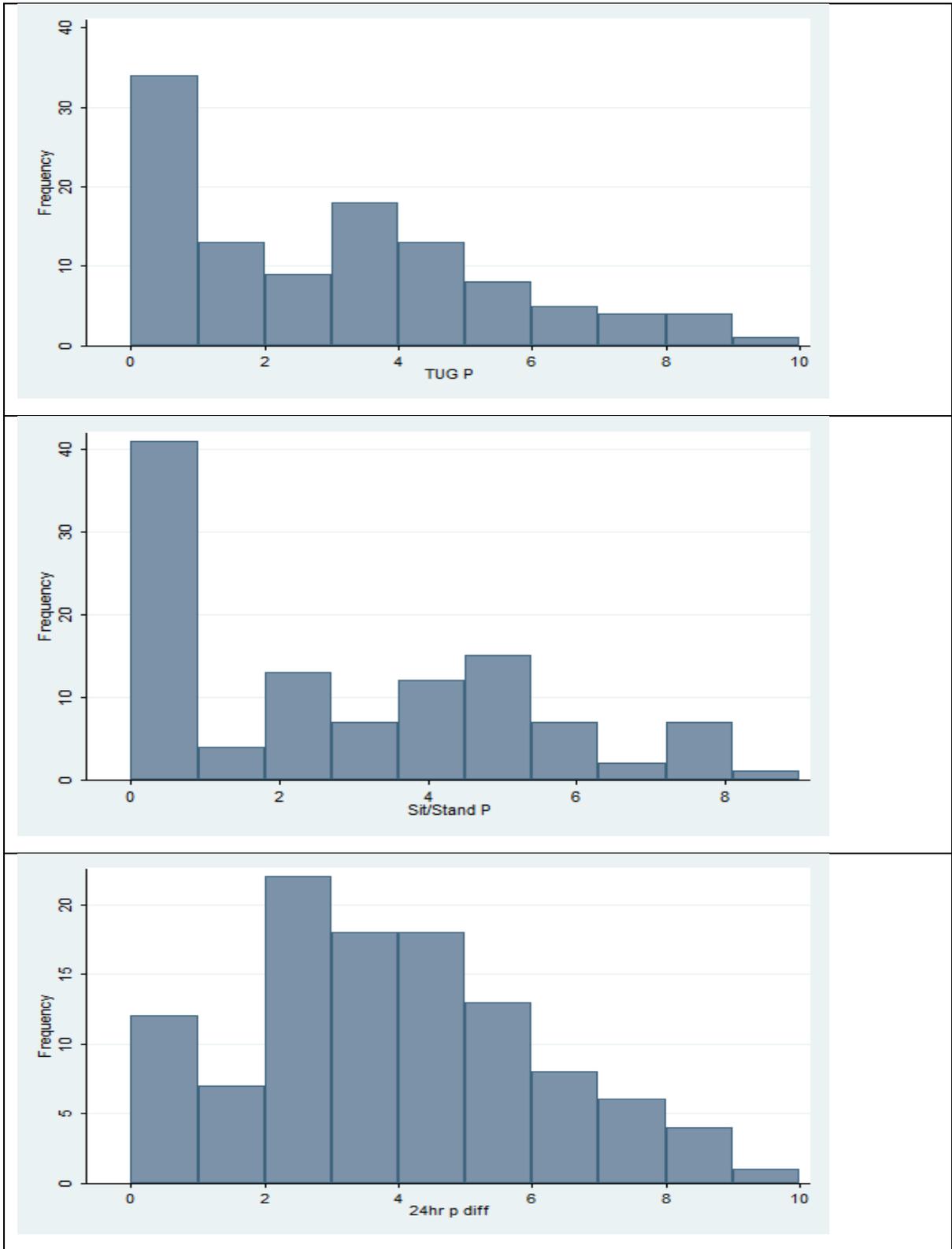


Figure 3.3 Distribution of pain scores from physical performance measures and change in pain (NPRS) over 24-hours

Analysis showed the i-score correlated moderately with pain reported with TUG ($r = 0.45$) and thirty second sit to stand ($r = 0.39$). Both results were significant ($p < 0.0001$). There was a weak correlation with pain difference over a twenty-four hour period ($r = 0.27$). This result was significant ($p < 0.005$).

Inter-rater reliability

Twenty-eight participants with knee OA provided data for the assessment of inter-rater reliability of the i-score (Appendix G). Analysis showed an ICC (2,1) of 0.64 (95% CI: 0.36, 0.82) for the i-score. This represents a moderate level of agreement (Portney and Watkins, 2000). SEM was 2.57 meaning the standard error of the scale, where scores could range between 4 points and 40 points, was 2.57 points. When using the scale a physiotherapist can be confident that the true score for irritability is likely to be the points generated by the i-score ± 2.57 .

Internal consistency

Cronbach's alpha (α) for the i-score was 0.59 which did not reach the 0.7 threshold indicative of good internal consistency (Nunnally, 1978).

Summary

A summary of results of testing psychometric properties of the i-score is shown in Table 3.6.

Table 3.6 Psychometric properties of the i-score

Property	Test (Sample size)	Result
Content validity	NGT (n = 4)	4 out of 5 top ranked items by NGT mapped to the 4 items included in the i-score
Criterion validity	Pearson correlation (n = 83)	r = 0.48* (the i-score and physiotherapists existing method of assessment)
Construct validity	Pearson correlation (n = 109)	r = 0.45* (the i-score and TUG pain) r = 0.39* (the i-score and SST pain) r = 0.27† (the i-score and PDIFF in 24-hours)
Inter-rater reliability	ICC (2,1); SEM (n = 28)	ICC (2,1)= 0.64 (95%CI: 0.36, 0.82) SEM = 2.57
Internal consistency	Cronbach's alpha (α)	$\alpha = 0.59$

* p < 0.0001; † p < 0.005

NGT = nominal group technique; TUG pain = pain reported with timed up and go test; SST pain = pain reported with 30-second sit to stand test; PDIFF = difference in pain in 24-hour period;

ICC – intraclass correlation coefficient; SEM = standard error of measurement

3.6. Discussion

The i-score is a new scale developed to assess and report musculoskeletal irritability in individuals with knee or hip OA. It has a minimum score of four and a maximum score of forty, with higher scores indicating greater irritability. The i-score provides a standardised process with clear operational definitions (Tables 3.1 and 3.3). Additionally, investigation of its psychometric properties demonstrates acceptable validity and reliability.

3.6.1. Previous research

Items in the i-score use the same components as a study exploring Maitland's concept of irritability in patients with low back pain (LBP) (Barakatt et al., 2009b). However, the previous study, using principal component analysis, did not identify ease of aggravation (time to onset in the i-score) or pain severity (NPRS in the i-score) as important for grouping judgements of irritability; whereas pain persistence (time to pain relief in the i-score) and aggravating activity were statistically important (Barakatt et al., 2009b). A point of difference is the i-score separated aggravating activity so that each item measured a discrete construct. Beyond this, while findings from the previous study suggest some factors might be more important than others, a deliberate decision was made not to weight items in the i-score, in accordance with recommendations for scale development (Streiner and Norman, 2003).

3.6.2. Strengths

The i-score was developed using robust and recommended methods (Streiner and Norman, 2003; DeWalt et al., 2007); a claim supported by the results of psychometric testing. With a structure that incorporated reported components of irritability, the resulting scale is reflective of clinical practice, which should increase its clinical utility and lessen potential barriers to implementation.

Content validity was demonstrated using NGT, which proved a useful technique for gaining consensus from physiotherapists involved in the care of knee and hip OA participants, on the important factors for assessing irritability. The investigation was easy to perform, in a short space of time and consensus was achieved relatively easily.

Evidence of criterion validity of the i-score was provided by moderate correlation ($r = 0.48$) with physiotherapists' existing method of irritability assessment. Similarly, positive correlations with related measures of pain behaviour with activity (PPMs) provide evidence of construct validity. However, no existing measures comprehensively assessed irritability, as indicated by moderate rather than high correlation values. These results justify the development of the new tool.

In clinical practice, where different physiotherapists might treat the same patient, good inter-rater reliability of the i-score means that a change greater than 2.6 (SEM) on the scale would reflect actual change in irritability, rather than error of assessment. Additionally, ICC (2,1) of 0.64 is interpreted as moderate reliability (Portney and Watkins, 2000). Same day testing gave the advantage of knowing level of irritability was the same for both testers.

Internal consistency ($\alpha = 0.59$) was slightly lower than is considered ideal (Nunnally, 1978). This could be due in part to there being only four items in the scale, as it is recognised that scales with more items tend to have higher α scores (Streiner and Norman, 2003). An alternative explanation is that the i-score includes diverse aspects of pain behaviour which aim to improve content validity, and these do not correlate strongly with each other. Internal consistency could be improved by removing items with low correlation but this would be at the cost of decreasing content validity, which is arguably more important for this scale (Streiner and Norman, 2003).

3.6.3. Limitations

A problem in the development of the i-score was the lack of previous research into irritability, and also the domains measured by the four items in the scale. A particular issue was grading aggravating activities in terms of ease or difficulty for patients with hip or knee OA, which appears to be a different approach. Previous assessment of function has relied on patient report (e.g. WOMAC) rather than consideration of biomechanical forces, speed of movement, or duration of activity. This approach therefore needs further research to determine its validity. Additionally, no instruction was given regarding level of detail required for the aggravating activity item. This resulted in some reported activities being hard to categorize.

Other limitations concerned time categories for onset and easing of symptoms, which were based on previous experience of patient reports. Division into four categories for each was problematic, with some difficulty equating time categories to the other two items. This resulted in the least severe categories (unlimited time to pain onset, and less than one minute to pain relief) being assigned two points each, giving a minimum i-score of four.

Limitations of the psychometric testing include: the use of only four physiotherapists, fewer than the recommended eight to twelve participants for NGT, which may have jeopardised the reliability of the findings (Black, 2006). All of the physiotherapists were involved in the clinical trial and at other times worked alongside each other in the same clinic. It is possible they had formed similar views regarding patient management generally. Furthermore they had been directed to record patient irritability as part of the standardised intervention for the clinical trial, which may have sensitised them to irritability assessment more than usual. All these issues could limit the generalizability of results.

For criterion validity, a short period of time elapsed between baseline assessment when the i-score was calculated, and the first treatment session (up to three weeks), so it is feasible there was a change in patient irritability which could have affected the correlation value for criterion validity. Conversely, same day testing may have increased participant recall of response, and inflated results for inter-rater reliability.

3.6.4. Future research

Future research could investigate adding additional items to the i-score to improve content validity further, given the range of factors identified in NGT. If number of items increased, internal consistency may increase, especially if additional items correlated with each other or the total score. NGT offers a good method for undertaking future research to confirm preliminary findings for content validity. In addition to building on evidence for use of the i-score in knee and hip OA patients, successful use of the technique for this study provides support for use of NGT to examine similar topics.

The time periods selected for time to onset or time to easing resulted in irregular intervals. They could be refined by patients recording actual time to onset or easing of symptoms rather than relying on recall.

Future versions of the i-score should seek to increase inter-rater reliability by providing clear operational instructions for the conduct of the questioning.

Future development of the scale for other conditions would necessitate identifying and categorizing aggravating activities for each specific condition. This would involve developing multiple different i-scores, which would be time-consuming and costly; in addition to over-burdening clinicians with numerous tools. An alternative may be to remove the aggravating activity item from the scoring system. Instead a typical activity could be nominated and recorded. The remaining three items could be scored as described for the i-score. Providing

other psychometric properties of the i-score were maintained or improved, this would move it from being a condition specific measure to a dimension specific (irritability) measure (Hefford et al., 2011), and arguably increase its clinical utility.

It can be seen that construction of a scale is a balance between producing a simple tool that captures the underlying construct, and production of a more complex tool that has robust psychometric properties, but may be more cumbersome to implement. Item response theory (IRT) provides an alternative approach to classical test theory (CTT), addressing many of its shortcomings and facilitating development of shorter scales without any loss of reliability (Streiner, 2010). Use of IRT and related methods of factor analysis could further ongoing development of the i-score.

3.6.5. Implications for use of the i-score

The concept of irritability is one that is familiar to physiotherapists. It contributes to the clinical reasoning process, and should be considered when adapting clinical guidelines or applying clinical decision aids to the individual patient. Experienced clinical practitioners have an intuitive feel for assessing irritability (Smart and Doody, 2007); however, standardising the process should enhance communication between therapists involved in the care of an individual, thereby improving patient care. Provision of a framework that standardises assessment and reporting of irritability should facilitate teaching of students or novice practitioners, leading to better understanding of the concept and application to clinical practice. Additionally the i-score has evidence of content, criterion and construct validity and is able to provide a consistent and reliable measure for research purposes.

3.7. Conclusions

Development of the i-score was an innovative approach to standardising the assessment and reporting of irritability. It met the requirements for use in the prognostic study, enabling irritability to be investigated as a possible predictor of outcome following physiotherapy intervention for patients with knee OA.

4. Inter rater reliability of baseline measures

4.1. Introduction

Reliability is the extent to which a measure can be relied upon to be consistent across repeated measures, and free from measurement error (Portney and Watkins, 2000). Measurement error affecting reliability can arise from a number of sources including changes in the testing environment, the accuracy or reliability of testing equipment, or changes in the clinical condition of the patient (Portney and Watkins, 2000). Reliability of clinical tests can also be affected by inter-rater reliability which looks at the agreement between two or more raters for the results of the same test. Good inter-rater reliability allows the researcher to assume that the measurements obtained by examiners are likely to be representative of the subject's true score, and therefore can be interpreted and applied with greater confidence to a larger population (Portney and Watkins, 2000). Inter-rater reliability is relevant to the study of predictive variables, as the value lies in many investigators being able to measure the variable of interest reliably and then using it to predict outcome. Consistent results are important if clinicians are to be confident making treatment decisions (Hicks et al., 2003). Inter-rater reliability is frequently reported in predictive studies and allows the reader to judge the threat it may pose to the internal validity and generalizability of the findings (Hayden et al., 2006; Beattie and Nelson, 2007).

An essential component of the current prognostic study was the construction of a standardised baseline clinical assessment for patients with knee osteoarthritis (OA). While inter-rater reliability was not the only consideration for inclusion of potential predictors, it is useful to be able to report the inter-rater reliability of included variables. The aim therefore of the reliability study was to identify inter-rater reliability of variables routinely used in baseline assessment of

patients with knee OA, including information from both patient history and physical examination.

Preliminary examination of the literature revealed a lack of reported inter-rater reliability for some variables in routine clinical use. It was therefore necessary to use two approaches to establish inter-rater reliability of baseline assessment clinical tests in an OA population. The first was to review the literature pertaining to reliability of clinical tests in patients with lower limb OA or in an older age group. Secondly, an investigation was conducted to determine inter-rater reliability for clinical tests where reliability had not previously been established or where results were inconclusive.

4.2. Methods

4.2.1. Literature review

Inter-rater reliability has previously been investigated and reported for a number of variables used to assess knee OA. Reliability studies were identified from reference lists in textbooks (Magee, 2002; Cleland, 2005) and by searching electronic databases (Medline, Embase, AMED, CINAHL). Studies were reviewed for relevance to the current investigation i.e. clinical tests of the knee and lower limb; study population consisted of older adults or those suffering from osteoarthritis; and tests in common clinical use, that do not require sophisticated techniques of measurement or equipment. Tests included variables related to patient history, as well as physical examination, as both sources of information are necessary to obtain a true clinical picture, and both are routinely used in clinical practice. For the purposes of this reliability study patient history referred to any information obtained via questionnaires, or oral interview. It included demographic details, psychosocial information, disease history, and symptoms. Physical examination referred to all information requiring an assessor/examiner/clinician to perform a physical test on the patient. It included

tests for alignment or positioning, joint movement, accessory movements, muscle strength, ligament and meniscal tests, palpation, and physical performance measures (PPM).

4.2.2. Investigation of inter-rater reliability for clinical tests

Reliability data were not available for all tests which could be considered for inclusion in the baseline clinical assessment of the prognostic study. These tests were the focus of the current inter-rater reliability investigation. Consistent with the aforementioned literature review, such tests addressed a number of domains considered relevant to the examination of knee OA. They also had to be in routine clinical use, and not require any sophisticated testing equipment.

The sample consisted of hip and knee OA patients drawn from participants in a clinical trial (Abbott et al., 2009). An extension of ethical approval for the Management of Osteoarthritis (MOA) clinical trial was granted by the Lower South Regional Ethics Committee of the New Zealand (Project Key: LRS/07/11/044) for the reliability study (Appendix I). Consecutive patients attending for the baseline assessment in the trial were invited to return for a repeat assessment of selected physical examination variables. All patients in the reliability study provided additional informed consent following an explanation of the purpose of the reliability study.

Sample size was calculated using a method designed for reliability studies (Walter et al., 1998). Based on a level of significance, $\alpha=0.05$ and a type II error, $\beta = 0.20$, sample size was estimated to be thirty-three subjects. This was for a test: retest study ($n = 2$ testers) with a minimal accepted level of reliability ($p_0 = 0.40$ or “fair”) and a desired level of reliability ($p_1 = 0.70$ or “substantial”) (Sim and Wright, 2005).

The two raters were physiotherapists who were the primary assessors for hip and knee OA in the clinical trial. A manual of operating procedures, detailing

performance of clinical tests, was developed for the main trial (Appendix K). This included definitions of patient positioning and interpretation of test results. Prior to the reliability study, the two raters undertook refresher training sessions to standardize their performance.

Participants underwent a full examination of either hip or knee, as part of the main trial, and measurements were recorded by rater one. A second appointment was made within seven days, prior to any treatment. It was important that the subjects were retested before they received any intervention that had potential to change the clinical features being measured. Upon return, the patient underwent additional testing by rater one so as to complete the clinical measures for the reliability assessment. Rater two then performed the whole reliability assessment, which included a combination of measures for both the hip and knee. The second rater was blinded to the results of the first rater.

The patient was examined in the same environment with the same equipment on both visits to reduce measurement error associated with external factors. A hand held dynamometer (Nicholas MMT, Model 01160, Lafayette Instruments, Lafayette, Indiana) and digital weighing scales were the only equipment used for assessment, and these were calibrated regularly. A copy of the reliability assessment can be seen in (Appendix J).

4.2.3. Data analysis

Inter-rater reliability was reported for non-ordinal categorical data with Cohen's kappa coefficient (to account for chance agreement) and percent agreement (Tooth and Ottenbacher, 2004; Sim and Wright, 2005). Type 2,1 intraclass correlation coefficients (ICC) were used for continuous and ordinal variables, with corresponding standard errors of measurement (SEM) (Portney and Watkins, 2000). As the subjects were assessed by two raters representative of the usual population of physiotherapists and the results were single scores (not

means), ICC model 2,1 was used (Portney and Watkins, 2000). The SEM quantifies the measurement error in the same units as the original measurement providing a clinically meaningful value to be interpreted by the clinician (Stratford and Goldsmith, 1997). Ninety-five percent confidence intervals (95% CIs) were calculated for all reliability coefficients.

The Bland-Altman approach is also recommended by some authors to examine the difference between paired readings, in this case the difference in results of clinical tests performed by two examiners (Bland and Altman, 1986; Bland and Altman, 1999; Petrie, 2006). It is used when data are continuous. It has the advantage of using the actual units of measurement (e.g. degrees of movement), providing meaningful information that can be easily interpreted by clinicians. Bland-Altman plots display mean value of rater differences with the upper and lower limits of agreement; the interval between them contains approximately 95% of the rater differences (Petrie, 2006). One of the assumptions underlying the Bland-Altman approach is that the differences between raters must be normally distributed. Normality was tested statistically with the Shapiro-Wilks test for normality. A p-value <0.05 meant the data was not normally distributed and further analysis was not performed for that variable. When p values of the Shapiro-Wilks test were approaching 0.05 the data was further examined with histograms and q-q plots to determine normal distribution. The following qualitative interpretations for kappa coefficients were used: excellent = 0.81 to 1.00; substantial = 0.61 to 0.80; moderate = 0.41 to 0.60; fair = 0.21 to 0.40; slight = 0.00 to 0.20; and poor = < 0.00 (Landis and Koch, 1977). Qualitative interpretations for ICC (2,1) values used in this study were: ICC values > 0.75 indicate good reliability; values from 0.50 to 0.75 are indicative of moderate reliability; and values < 0.50 are indicative of poor reliability (Portney and Watkins, 2000).

All statistical analyses were performed using STATA 10.1 (StataCorp, 4905 Lakeway Drive, College Station, Texas 77845, USA).

4.3. Results

4.3.1. Literature review

Variables from patient history.

Inter-rater reliability values from previous studies are shown in Table 4.1

Depression was evaluated using the two-item case-finding instrument for detection of depression (Whooley et al., 1997). Self-efficacy, fear avoidance, and catastrophizing were evaluated using the Pain Belief Screening Instrument (PBSI) (Sandborgh et al., 2007). Both instruments have proven validity and are able to classify patients with accuracy (Whooley et al., 1997; Sandborgh et al., 2007; Sandborgh et al., 2008). Neither reported inter-rater reliability nor were any studies found in which these tools were applied to a knee OA population.

With the exception of these two measures of depression and pain beliefs, the majority of variables relating to the patient history and symptoms had at least moderate evidence of inter-rater reliability.

Table 4.1 Inter-rater reliability of patient history variables from review of literature

Variable	Reference	Inter-rater reliability	Population/Study details	Level of agreement*
Disease history				
Previous knee injury	(Peat et al., 2003)	% agreement = 91.3 $\kappa = 0.62$	n = 58 age > 50 years, knee pain	Substantial
Use of walking aid	(Peat et al., 2003)	% agreement = 93.1 $\kappa = 0.85$	n = 58 age > 50 years, knee pain	Excellent
Symptoms				
NPRS	(Mawdsley et al., 2002)	All subjects: ICC 0.74, SEM 0.66 Subjects with no change in pain: ICC 0.90, SEM 0.68	n = 32 age > 61 years, general musculoskeletal subjects	Moderate (all subjects) Substantial (subjects with no change in pain)
Unilateral v bilateral pain	(Peat et al., 2003)	% agreement = 86.2 $\kappa = 0.72$	n = 58 age > 50 years, knee pain	Substantial
Current hip problems	(Peat et al., 2003)	Right hip: % agreement = 93.0 $\kappa = 0.71$ Left hip: % agreement = 87.7 $\kappa = 0.56$	n = 58 age > 50 years, knee pain	Substantial (right hip) Moderate (left hip)

Variable	Reference	Inter-rater reliability	Population/Study details	Level of agreement*
Knee locking	(Peat et al., 2003)	% agreement = 91.4 $\kappa = 0.71$	n = 58 median age = 63 years (50-86) knee pain	(Peat et al., 2003) Substantial
	(Dervin et al., 2001)	% agreement = 80 $\kappa = 0.44$ (95%CI: 0.26, 0.62)	n = 115 age 60.5 years (± 8.5 years) knee OA patients	(Dervin et al., 2001) Moderate
Knee giving way/instability (actual or feeling of)	(Peat et al., 2003)	% agreement = 87.7–89.5 $\kappa = 0.73 - 0.77$	n = 58 median age = 63 years (50-86) knee pain	(Peat et al., 2003) Substantial
	(Dervin et al., 2001)	% agreement = 60 $\kappa = 0.12$ (95%CI: -0.04, 0.28)	n = 115 age 60.5 years (± 8.5 years) knee OA patients	(Dervin et al., 2001) Slight
Self-report of instability from Knee Outcome Survey-Activities of Daily Living Scale (KOS-ADLS) (Irrgang et al., 1998)	(Fitzgerald et al., 2004a)	ICC 0.72	n = 50 Knee patients including OA	Moderate

Variable	Reference	Inter-rater reliability	Population/Study details	Level of agreement*
Knee swelling	(Peat et al., 2003)	% agreement = 79.3 $\kappa = 0.57$	n = 58 median age = 63 years (50-86) knee pain	(Peat et al., 2003) Moderate
	(Dervin et al., 2001)	% agreement = 69 $\kappa = 0.33$ (95%CI: 0.17, 0.49)	n = 115 age 60.5 years (± 8.5 years) knee OA patients	(Dervin et al., 2001) Fair
Irritability	Thesis chapter	ICC (2,1) 0.643 (95%CI: 0.361, 0.817) SEM = 2.57	n = 28 age = 67 years (37 – 92), knee OA patients	Moderate
Disturbed sleep	(Peat et al., 2003)	% agreement = 86.2 $\kappa = 0.79$	n = 58 median age = 63 years (50-86) knee pain	Substantial
ACR clinical criteria				
Knee pain (most days in last month)	(Peat et al., 2003)	% agreement = 94.7 $\kappa = 0.86$	n = 58 median age = 63 years (50-86) knee pain	Excellent

Variable	Reference	Inter-rater reliability	Population/Study details	Level of agreement*
Morning stiffness ≤ 30min	(Peat et al., 2003)	% agreement = 98.3 κ = 0.79	n = 58 median age = 63 years (50-86) knee pain	(Peat et al., 2003) Substantial
	(Jones et al., 1992)	κ = 0.58 (95%CI: 0.38, 0.79)	n = 49 (98 knees) age 50 -92, 45 elderly in-patients, 4 rheumatology out-patients knee OA	(Jones et al., 1992) Moderate
<p>* level of agreement: ICC > 0.75 = good, 0.5 – 0.75 = moderate, < 0.5 = poor (Portney and Watkins, 2000); Kappa ≥ 0.81 – 1 = excellent, 0.61 – 0.8 = substantial, 0.41 – 0.6 = moderate, 0.21 – 0.4 = fair, 0.0 – 0.2 = slight, < 0.0 = poor (Landis and Koch, 1977); κ = kappa statistic; NPRS = numeric pain rating scale; ICC = intraclass correlation coefficient; SEM = standard error of measurement; 95%CI = 95% confidence interval; OA = osteoarthritis; ACR = American College of Rheumatology</p>				

Physical examination variables

Inter-rater reliability values from previous studies are shown in Table 4.2. In summary approximately half of the selected physical examination variables have at least moderate inter-rater reliability, as reported in the literature.

Table 4.2 Inter-rater reliability of physical examination variables from literature review

Variable	Reference	Inter-rater reliability	Population/Study details	Level of Agreement *
Joint alignment/position				
Knee valgus/varus deformity	(Wood et al., 2006)	<i>Valgus</i> % agreement = 63.2 $\kappa = 0.11$ (lower 99%CL, -0.20) <i>Varus</i> % agreement = 77.2 $\kappa = 0.41$ (lower 99%CL, 0.11)	n = 57 median age = 63 years (50-86) knee pain	Valgus – Slight Varus – Moderate
Medial-lateral positioning of patella	(Fitzgerald and McClure, 1995) (Herrington, 2000)	% agreement = 44 $\kappa = 0.10$ <i>Medial distance</i> ICC = 0.91 <i>Lateral distance</i> ICC = 0.94	n = 66 age = 14 – 74 years patello-femoral pain n = 1 (20 raters) “normal subject”	(Fitzgerald and McClure, 1995) Slight (Herrington, 2000) Good
Longitudinal arch angle	(Jonson and Gross, 1997)	ICC (2,1) 0.81 Mean absolute difference = 5°	n = 63 age 18 – 30 years healthy male recruits	Excellent
Fixed flexion deformity	(Wood et al., 2006)	% agreement = 80.4 $\kappa = 0.53$	n = 56 median age = 63 years	Moderate

Variable	Reference	Inter-rater reliability	Population/Study details	Level of Agreement *
		(lower 99%CL, 0.24)	(50-86) knee pain	
Joint Movement: ROM, end feel, pain/resistance sequence, muscle length				
Knee range of motion (goniometer)	(Watkins et al., 1991)	<i>Knee flexion</i> ICC (1,1) = 0.90	n = 43 (50 knees) age = 39.5 years (± 15) patients with range of conditions	Good (all measures/ all studies)
	(Fritz et al., 1998)	<i>Knee extension</i> ICC (1,1) = 0.86 <i>Knee flexion</i> ICC (2,1) 0.97 SEM 3.9 degrees	n = 35 age = 40 years (± 15.9) patients with knee dysfunction	
	(Brosseau et al., 2001)	<i>Knee extension</i> ICC (2,1) 0.94 SEM 1.7 degrees <i>Knee flexion</i> ICC (2,1) 0.977 – 0.982 <i>Knee extension</i> ICC (2,1) 0.893 – 0.926	n = 60 mean age = 52 patients with knee restrictions (including arthritis = 31)	
Pain-resistance sequence	(Fritz et al., 1998)	<i>Knee flexion</i> % agreement = 74 κ = 0.28 (weighted)	n = 35 age = 40 years (± 15.9) unilateral knee dysfunction	(Fritz et al., 1998) Fair
	(Hayes and Petersen, 2001)	<i>Knee flexion</i> % agreement = 52.9 κ = 0.51	n = 17 age = 31.8 years (± 9.5) unilateral knee pain	(Hayes and Petersen, 2001) Moderate

Variable	Reference	Inter-rater reliability	Population/Study details	Level of Agreement *
		<i>Knee extension</i> % agreement = 58.5 $\kappa = 0.42$		
End feel	(Hayes and Petersen, 2001)	<i>Knee flexion</i> % agreement = 35.3 $\kappa = -0.01$ (95%CI: -0.36, 0.35) <i>Knee extension</i> % agreement = 76.4 $\kappa = 0.43$ (95%CI: -0.06, 0.92)	n = 17 age = 31.8 years (± 9.5) unilateral knee pain	Flexion – Poor Extension - Moderate
Concordant pain	(Dervin et al., 2001)	<i>Full flexion</i> % agreement = 67 $\kappa = 0.09$ (95%CI: - 0.07, 0.25) <i>Full extension</i> % agreement = 57 $\kappa = 0.10$ (95%CI: - 0.08, 0.28)	n = 115 age = 60.5 years (± 8.5) knee OA patients	Slight
Hip ROM	(Lin et al., 2001)	<i>Right hip flexion</i> ICC 0.82 (95%CI: 0.26, 0.95) SEM = 4.34° <i>Left hip flexion</i>	n = 106 age = 69.4 years (± 5.9) 52 knee OA patients 17 hip OA patients 37 knee & hip OA	(Lin et al., 2001) Good

Variable	Reference	Inter-rater reliability	Population/Study details	Level of Agreement *
	(Cibere et al., 2008)	ICC 0.83 (95%CI: 0.33, 0.96) SEM = 3.64° <i>Hip flexion</i> R _c = 0.91 <i>Hip extension</i> R _c = 0.66 <i>Hip abduction</i> R _c = 0.88 <i>Hip adduction</i> R _c = 0.56	patients n = 6 median age = 63 years (49 -65) mild to severe radiographic hip OA	(Cibere et al., 2008) <i>Hip flexion & abduction</i> – adequate <i>Hip extension & adduction</i> – not adequate
Ankle ROM	(Elveru et al., 1988)	<i>Ankle dorsiflexion</i> ICC(1,1) 0.50	n = 49 age = 35.9 years (±15.6) neurological & orthopaedic disorders	Moderate
Thomas test for hip flexor tightness	(Clapis et al., 2008) (Cibere et al., 2008)	ICC (3,2) 0.92 PABAK = 0.88 Prevalence = 0.36 (post-standardisation)	n = 42 age = 22 years (± 3.71) healthy, no lower limb dysfunction n = 6 median age = 63 years (range 49 -65 years) mild to severe radiographic hip OA	(Clapis et al., 2008) Good (Cibere et al., 2008) Excellent

Variable	Reference	Inter-rater reliability	Population/Study details	Level of Agreement *
90/90 test for hamstring tightness	(Sutlive et al., 2004)	$\kappa = 0.19$	n = 30 age = 18 – 40 years active duty military population with patello-femoral pain syndrome (PFPS)	Slight
Muscle strength with hand held dynamometers (HHD)				
Knee extensors	(Dunn and Iversen, 2003)	<i>Right knee</i> ICC 0.92 (95%CI: 0.83, 0.96) <i>Left knee</i> ICC 0.87 (95%CI: 0.72, 0.94)	Nicholas HHD n = 25 age 68.2 years (± 7.5) Elderly, chronic low back pain	Good
Knee flexors	(Dunn and Iversen, 2003)	<i>Right knee</i> ICC 0.91 (95%CI: 0.81, 0.96) <i>Left knee</i> ICC 0.93 (95%CI: 0.84, 0.97)	Nicholas HHD n = 25 age 68.2 years (± 7.5) Elderly, chronic low back pain	Good
Ligament/Meniscus tests				
McMurray/valgus/varus instability	(Dervin et al., 2001)	<i>McMurray</i> % agreement = 59 $\kappa = 0.16$ (95%CI: - 0.01, 0.33)	n = 115 age 60.5 years (± 8.5 years) knee OA patients	(Dervin et al., 2001) Slight

Variable	Reference	Inter-rater reliability	Population/Study details	Level of Agreement *
	(Wood et al., 2006)	<i>Valgus instability</i> % agreement = 92 $\kappa = 0.05$ (95%CI: - 0.03, 0.23) <i>Varus instability</i> % agreement = 93 $\kappa = 0$ (95%CI: - 0.18, 0.18) <i>Valgus stress</i> % agreement = 67.9 $\kappa = 0.36$ (lower 99%CL, 0.07) <i>Varus stress</i> % agreement = 76.8 $\kappa = 0.24$ (lower 99%CL, -0.12)	n = 56 median age = 63 years (range 50-86) knee pain	(Wood et al., 2006) Fair
Palpation tests				
Joint line tenderness	(Dervin et al., 2001)	<i>Medial joint line</i> % agreement = 79 $\kappa = 0.21$ (95%CI: 0.01, 0.41) <i>Lateral joint line</i> % agreement = 70 $\kappa = 0.25$	n = 115 age 60.5 years (± 8.5) knee OA patients	(Dervin et al., 2001) Fair

Variable	Reference	Inter-rater reliability	Population/Study details	Level of Agreement *
	(Jones et al., 1992)	(95%CI: 0.07, 0.43) <i>Medial tibiofemoral tenderness</i> $\kappa = 0.35$	n = 49 (98 knees) age 50 -92, 45 elderly in-patients, 4 rheumatology out-patients knee OA	(Jones et al., 1992) Fair
	(Wood et al., 2006)	(95%CI: 0.24, 0.45) <i>Lateral tibiofemoral tenderness</i> $\kappa = 0.29$ (95%CI: 0.14, 0.44) <i>Patellofemoral tenderness</i> $\kappa = 0.27$ (95%CI: 0.05, 0.48) <i>Medial tibiofemoral tenderness</i> % agreement = 64.9 $\kappa = 0.26$ (lower 99%CL, -0.04) <i>Lateral tibiofemoral tenderness</i> % agreement = 68.4 $\kappa = 0.29$ (lower 99%CL, -0.02)	n = 57 median age = 63 years (range 50-86) knee pain	(Wood et al., 2006) Fair
Swelling	(Dervin et al., 2001)	<i>Sweep test</i> % agreement = 58	n = 115 age 60.5 years (± 8.5)	(Dervin et al., 2001) Slight

Variable	Reference	Inter-rater reliability	Population/Study details	Level of Agreement *
	(Wood et al., 2006)	$\kappa = 0.14$ (95%CI: - 0.04, 0.32) <i>Bulge sign</i> % agreement = 57.9 $\kappa = -0.04$ (lower 99%CL, -0.33) <i>Patellar tap</i> % agreement = 96.5	knee OA patients n = 57 median age = 63 years (50-86) knee pain	(Wood et al., 2006) Poor
	(Fritz et al., 1998)	$\kappa = -0.02$ (lower 99%CL, -0.05) <i>Fluctuation test</i> % agreement = 71 $\kappa = 0.37$ <i>Patellar tap</i> % agreement = 71 $\kappa = 0.21$	n = 35 age = 40 years (± 15.9) patients with knee dysfunction	(Fritz et al., 1998) Fair
Warmth	(Jones et al., 1992)	$\kappa = 0.23$ (95%CI: 0, 1.00)	n = 49 (98 knees) age 50 -92, 45 elderly in-patients, 4 rheumatology out- patients knee OA	(Jones et al., 1992) Fair
	(Wood et al., 2006)	% agreement = 70.6 $\kappa = 0.18$ (lower 99%CL, -0.25)	n = 34 median age = 63 years (50-86)	(Wood et al., 2006) Slight

Variable	Reference	Inter-rater reliability	Population/Study details	Level of Agreement *
	(Fritz et al., 1998)	% agreement = 83 $\kappa = 0.66$	knee pain n = 35 age = 40 years (± 15.9) patients with knee dysfunction	(Fritz et al., 1998) Substantial
Crepitus	(Jones et al., 1992)	$\kappa = 0.23$ (95%CI: 0.01, 0.45)	n = 49 (98 knees) age 50 -92, 45 in-patients, 4 rheumatology out-patients knee OA	(Jones et al., 1992) Fair
	(Wood et al., 2006)	% agreement = 60.7 $\kappa = 0.22$ (lower 99%CL, -0.08)	n = 56 median age = 63 years (50-86) knee pain	(Wood et al., 2006) Fair
Patellofemoral compression	(Wood et al., 2006)	% agreement = 71.9 $\kappa = 0.43$ (lower 99%CL, 0.16)	n = 57 median age = 63 years (50-86) knee pain	Moderate
Bony enlargement	(Wood et al., 2006)	<i>Medial</i> % agreement = 72.7 $\kappa = 0.44$ (lower 99%CL, 0.15) <i>Lateral</i> % agreement = 54.6	n = 55 median age = 63 years (50-86) knee pain	Medial – Moderate Lateral - Slight

Variable	Reference	Inter-rater reliability	Population/Study details	Level of Agreement *
		$\kappa = 0.08$ (lower 99%CL, -0.23)		
Physical Performance Measures (PPM)				
Timed up and go (TUG)	(Kennedy et al., 2005)*	ICC (2,1) 0.75 (95%CI: 0.51, 0.89) SEM 1.07 sec (95%CI: 0.86, 1.41)	n = 21 age = 63.7 years (± 10.7) Knee and hip OA patients waiting for joint replacement surgery	Good
Forty metre self-paced walk (SPW)	(Kennedy et al., 2005)	ICC (2,1) 0.91 (95%CI: 0.81, 0.97) SEM 1.73 sec (95%CI: 1.39, 2.29)	n = 21 age = 63.7 years (± 10.7) Knee and hip OA patients waiting for joint replacement surgery	Good
Thirty-second chair stand test	(Gill and McBurney, 2008)	Baseline assessment: ICC (1,1) 0.93 (95%CI: 0.87, 0.96) SEM 0.7 stands	n = 42 age 70.3 years (± 9.8) Knee and hip OA patients waiting for joint replacement surgery	Good
Twenty-centimetre step-up test	(Mokkink et al., 2005)	DynaPort knee test – rise and descend cluster	n = 92 age = 67 years (± 9.9) 41 knee OA patients	Good

Variable	Reference	Inter-rater reliability	Population/Study details	Level of Agreement *
		ICC 0.85 (95%CI: 0.75, 0.91)	awaiting joint replacement surgery 51 knee OA patients post joint replacement	
<p>* level of agreement: ICC > 0.75 = good, 0.5 – 0.75 = moderate, < 0.5 = poor (Portney and Watkins, 2000); Kappa ≥ 0.81 – 1 = excellent, 0.61 – 0.8 = substantial, 0.41 – 0.6 = moderate, 0.21 – 0.4 = fair, 0.0 – 0.2 = slight, < 0.0 = poor (Landis and Koch, 1977); κ = kappa; 99%CL = 99% confidence limit; ICC = intraclass correlation coefficient; SEM = standard error of measurement; 95%CI = 95% confidence interval; OA = osteoarthritis; ROM = range of movement; R_c = reliability coefficient where > 0.8 chosen <i>a priori</i> to indicate adequate reliability(Cibere et al., 2008); PABAK = prevalence adjusted bias adjusted kappa; HHD = hand held dynamometer; † Test : retest not inter-rater</p>				

4.3.2. Physical examination – investigation of inter-rater reliability of clinical measures

Twenty-nine subjects completed the reliability study: fourteen with hip OA; fifteen with knee OA. There were seventeen female and twelve male participants; mean age was 69 years (range 46 to 84 years). Both hip and knee tests were performed on all patients. The main prognostic study is concerned with baseline assessment variables that predict outcome in knee OA patients, so only results of tests relevant to clinical assessment for knee OA participants are included here. (Full results in Appendix L).

Results for inter-rater reliability of clinical measures with continuous data are shown in Table 4.3. The ICC (2,1) values ranged from 0.18 to 0.95, with associated SEM values ranging from 0.72 to 12.2.

Results for inter-rater reliability of clinical measures with categorical data are shown in Table 4.4. Kappa (κ) values ranged from -0.12 to 0.66.

Table 4.3 Inter-rater reliability of clinical measures reported with ICC (2,1) in participants with knee or hip osteoarthritis

Variable	ICC (2,1) 95% CI	SEM 95%CI	Level of agreement *
Joint Movement: ROM, end feel, pain/resistance sequence, muscle length			
Hip flexion	0.80 (0.60, 0.91)	8.60 degrees	Good
Thomas Test, one-joint hip flexor restriction (psoas)	0.71 (0.47, 0.85)	6.69 degrees	Moderate
Thomas Test, two-joint hip flexor restriction (rectus femoris)	0.60 (0.31, 0.79)	12.2 degrees	Moderate
90/90 test, hamstring restriction	0.58 (0.28, 0.78)	7.58 degrees	Moderate
Muscle strength with hand held dynamometers (HHD)			
Hip flexors	0.74 (0.48, 0.87)	2.13 kg	Moderate
Knee extensors @ 90 degrees	0.72 (0.48, 0.86)	3.51 kg	Moderate
Hip external rotators	0.72 (0.48, 0.86)	1.23 kg	Moderate
Hip internal rotators	0.71 (0.47, 0.85)	1.28 kg	Moderate
Knee flexors @ 0 -20 degrees	0.68 (0.25, 0.86)	2.80 kg	Moderate
Hip abductors	0.65 (0.38, 0.82)	1.47 kg	Moderate
Ankle plantarflexors	0.40 (0.02, 0.68)	4.14 kg	Poor
Hip extensors	0.37 (-0.10, 0.70)	3.31 kg	Poor
Knee extensors @ 0 – 20 degrees	0.37 (-0.11, 0.71)	4.60 kg	Poor
Hip adductors	0.33 (-0.11, 0.68)	3.49 kg	Poor
Knee flexors @ 90 degrees	0.18 (-0.20, 0.51)	4.73 kg	Poor

Variable	ICC (2,1) 95% CI	SEM 95%CI	Level of agreement *
Physical Performance Measures (PPM)			
Forty metre self-paced walk (SPW)			
SPW time (seconds)	0.95 (0.90, 0.98)	2.00 sec	Good
SPW exertion (RPE)	0.79 (0.61, 0.90)	1.26	Good
SPW pain (NPRS)	0.84 (0.69, 0.92)	1.03	Good
Timed Up and Go (TUG)			
TUG time (seconds)	0.87 (0.74, 0.94)	0.84 sec	Good
TUG exertion (RPE)	0.92 (0.83, 0.96)	0.72	Good
TUG pain (NPRS)	0.76 (0.55, 0.88)	1.16	Good
Thirty second chair stand test (Sit – Stand)			
Sit – Stand (counted)	0.81 (0.63, 0.91)	1.27	Good
Sit – Stand exertion (RPE)	0.75 (0.53, 0.87)	1.33	Moderate
Sit – Stand pain (NPRS)	0.66 (0.39, 0.83)	1.49	Moderate
Step-up test			
Step-up test (counted)	0.91 (0.82, 0.96)	5.80	Good
Step-up test exertion (RPE)	0.88 (0.77, 0.94)	1.22	Good
Step-up test pain (NPRS)	0.88 (0.75, 0.94)	0.90	Good
ICC = intraclass correlation coefficient; 95%CI = 95% confidence interval; SEM = standard error of measurement; * level of agreement ICC > 0.75 = good, 0.5 – 0.75 = moderate, < 0.5 = poor (Portney and Watkins, 2000); ROM = range of movement; HHD = hand held dynamometry; RPE = rate of perceived exertion; NPRS = numerical pain rating scale			

Table 4.4 Inter-rater reliability of clinical variables reported with kappa (κ) in participants with knee or hip osteoarthritis.

Variable	Kappa (κ), 95% CI	% agreement	Level of agreement *
Joint alignment/position			
Longitudinal arch angle (LAA)	0.60 (0.35, 0.82)	86.2	Moderate
Knee valgus/varus alignment	0.30 (0.21, 0.64)	51.72	Fair
Fixed flexion deformity	0.21 (-0.25, 0.68)	55.17	Fair
Patellar position	0.16 (-0.32,0.64)	55.17	Slight
Joint Movement: End feel, pain/resistance sequence, concordant pain, accessory movement			
Hip flexion, concordant symptoms	0.66 (0.39, 0.93)	82.76	Substantial
Tibiofemoral P-A	0.63 (0.18, 1.0)	93.10	Substantial
Knee extension, end feel	0.59 (0.53, 0.76)	79.31	Moderate
Knee flexion, end feel	0.56 (0.52, 0.63)	75.86	Moderate
Knee flexion, pain resistance sequence	0.41 (0.28, 0.56)	62.07	Moderate
Talocrural A-P	0.41 (0.14, 0.55)	75.0	Moderate
Superior tibiofibular P-A	0.35 (0.27, 0.57)	62.07	Fair
Patellar caudal glide	0.29 (0.14, 0.58)	58.62	Fair
Knee extension, pain resistance sequence	0.28 (0.20, 0.39)	58.62	Fair
Talocrural P-A	0.26 (0.15, 0.32)	57.14	Fair
Superior tibiofibular A-P	0.22 (-0.12,0.48)	72.41	Fair
Patellar cephalic glide	0.19 (0.03, 0.28)	50.00	Slight

Variable	Kappa (κ), 95% CI	% agreement	Level of agreement *
Patellar medial glide	0.13 (-.02, 0.17)	62.70	Slight
Patellar lateral glide	0.10 (-0.06,0.22)	48.28	Slight
Tibiofemoral A-P	0.08 (0.01, 0.25)	58.62	Slight
Palpation tests			
Tenderness medial/lateral joint lines	0.60 (0.34, 0.86)	79.31	Moderate
Bony enlargement	0.58 (0.28, 0.87)	79.31	Moderate
Heat on palpation	0.46 (-0.17, 1.0)	93.10	Moderate
Patellar compression test	0.44 (0.03, 0.85)	82.76	Moderate
Tenderness patellar margins	0.37 (0.03, 0.71)	68.97	Fair
Tenderness greater trochanter	0.23 (-0.08,0.54)	62.07	Fair
Crepitus	0.22 (-0.13,0.57)	68.97	Fair
Wipe Test	-0.05 (-0.15,0.05)	89.67	Poor
* Level of agreement κ : 0.81 – 1 = excellent, 0.61 – 0.8 = substantial, 0.41 – 0.6 = moderate, 0.21 – 0.4 = fair, 0.0 – 0.2 = slight, < 0.0 = poor (Landis and Koch, 1977); P-A = postero – anterior accessory movement; A-P = antero – posterior accessory movement			

Positioning/ Malalignment

Tests for longitudinal arch angle (LAA) of the foot and alignment of the knee in the frontal plane have moderate and fair agreement, respectively. Tests for medial-lateral patella position show slight agreement, while tests for knee alignment in the sagittal plane show poor agreement.

Range of motion tests

All the variables examining joint range of motion indicated good or moderate agreement: hip flexion, Thomas test for one joint muscle tightness (iliopsoas), Thomas test for two joint muscle tightness (rectus femoris), and the 90/90 test for hamstring tightness. Only hip flexion range of motion data was normally distributed and therefore met assumptions for a Bland Altman plot. Actual values for mean difference between raters and limits of agreement are reported in degrees (Figure 4.1).

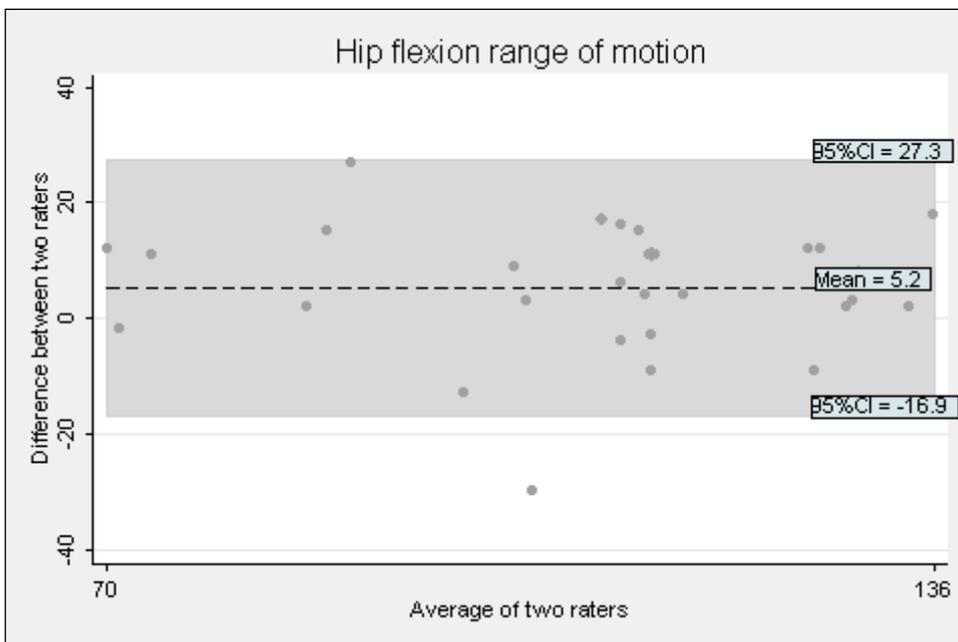


Figure 4.1 Bland Altman plot showing agreement between two raters for hip flexion range of motion (in degrees)

Accessory movements

Postero-anterior (P-A) accessory movement at the tibio-femoral joint show substantial agreement, and talocrural antero-posterior (A-P) show moderate agreement. Superior tibiofibular P-A, talocrural P-A and superior tibiofibular A-P showed fair agreement. One test, tibiofemoral A-P, showed slight agreement.

Muscle Strength Tests

Six muscle strength tests had moderate reliability: knee extension at ninety degrees, knee flexion at 0-20 degrees, hip internal rotation, hip external rotation, hip flexion, and hip abduction. Five had poor reliability: ankle plantarflexion, hip extension, knee extension at 0-20 degrees, knee flexion at ninety degrees, and hip adduction. Bland-Altman plots for strength tests with moderate reliability and normally distributed data are shown below (Figures 4.2. to 4.7). These display actual values for limits of agreement, reported in kilograms (kg), which can be easily interpreted for use in a clinical environment.

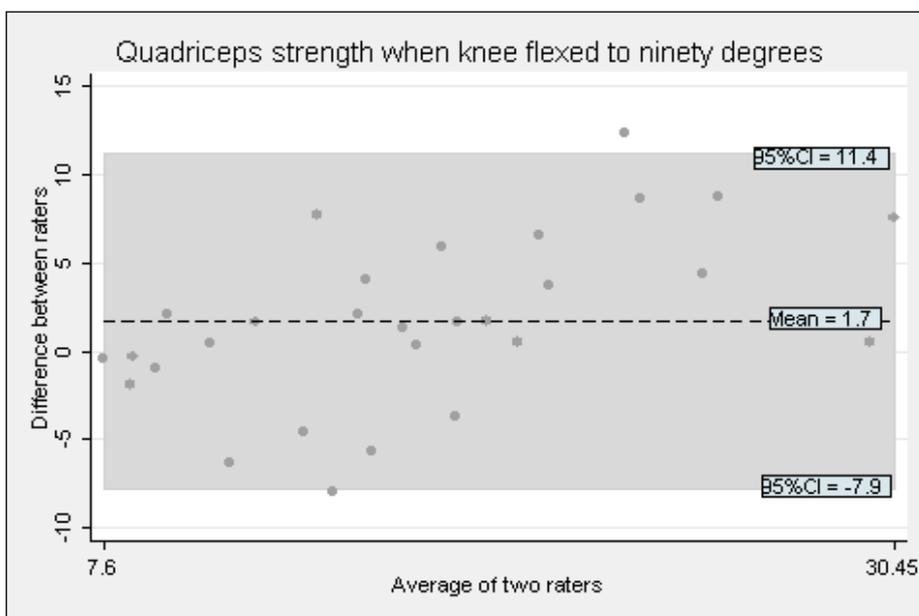


Figure 4.2 Bland Altman plot showing agreement between two raters for assessment of quadriceps strength “knee extension at ninety degrees” (in kg)



Figure 4.3 Bland Altman plot showing agreement between two raters for assessment of hamstrings strength “knee flexion at 0-20 degrees” (in kg)

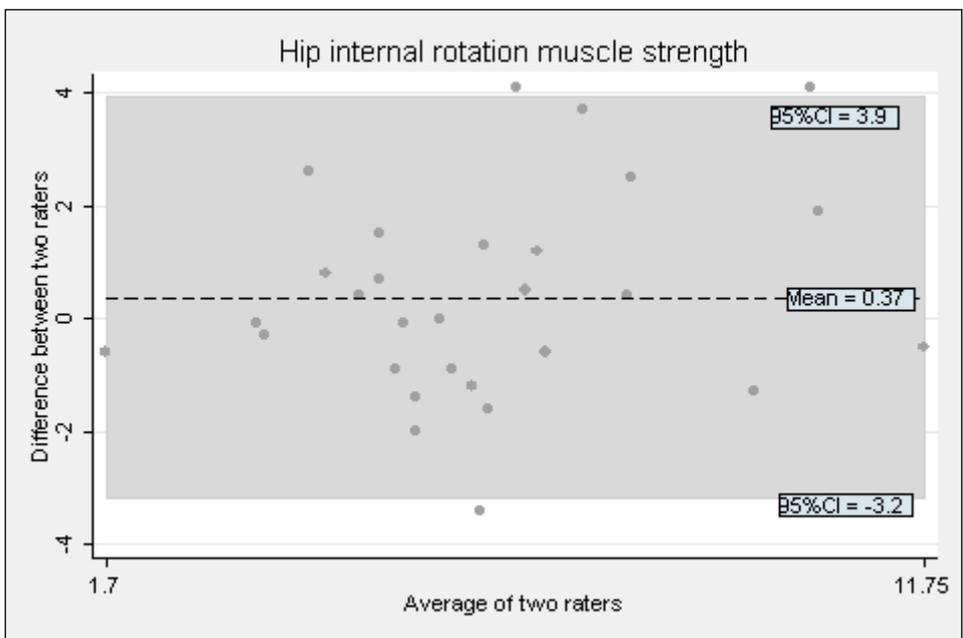


Figure 4.4 Bland Altman plot showing agreement between two raters for hip internal rotation muscle strength (in kg)

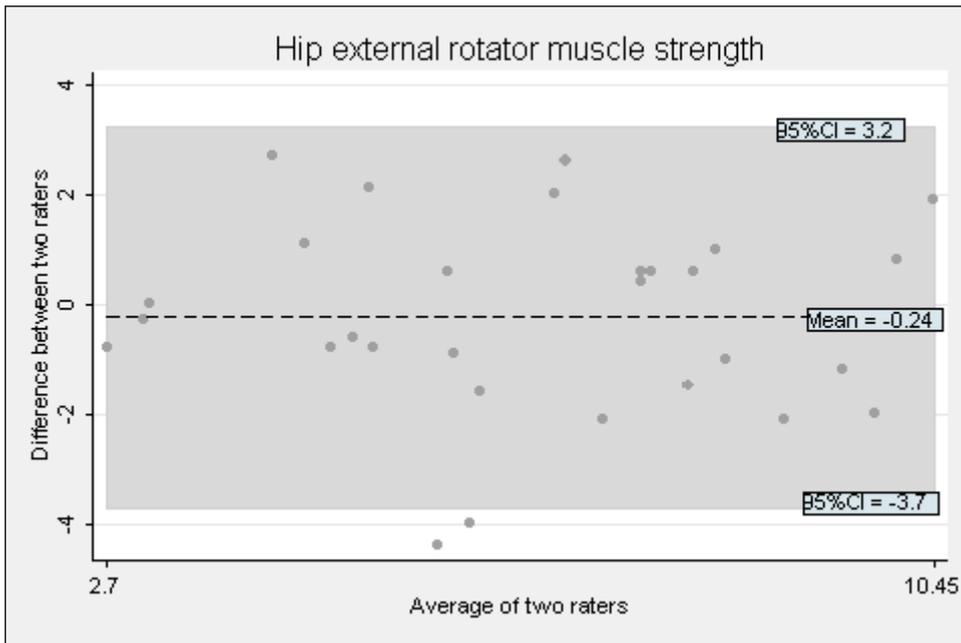


Figure 4.5 Bland Altman plot showing agreement between two raters for hip external rotation muscle strength (in kg)

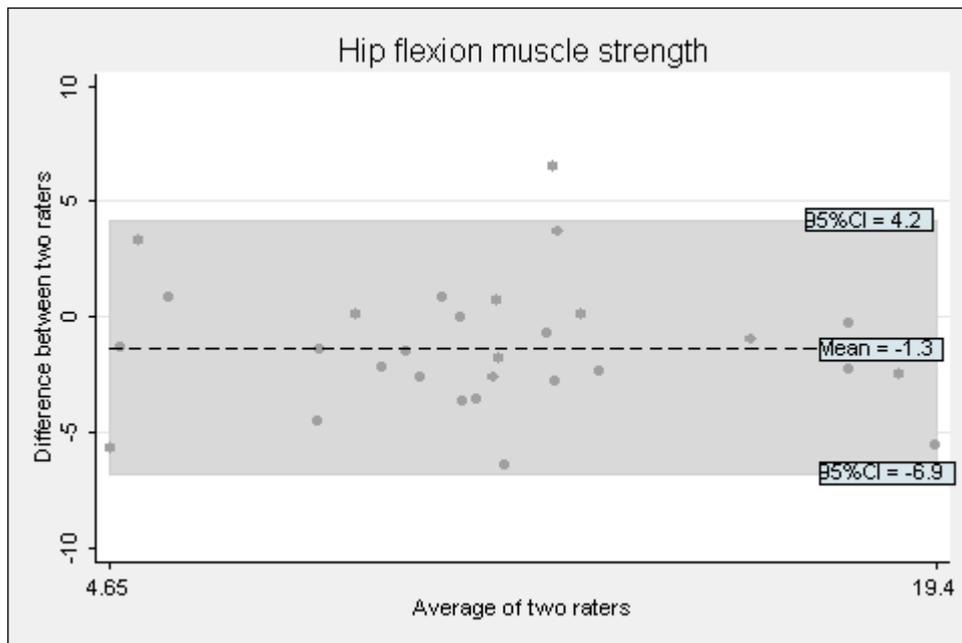


Figure 4.6 Bland Altman plot showing agreement between two raters for hip flexion muscle strength (in kg)

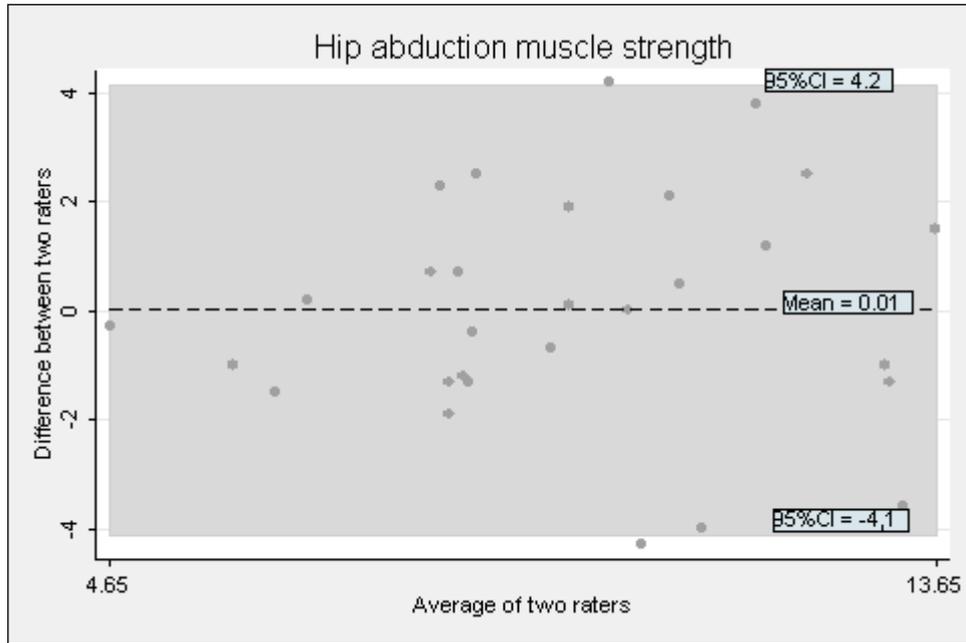


Figure 4.7 Bland Altman plot showing agreement between two raters for hip abduction muscle strength (in kg)

Pain/Symptom provocation tests

Inter-rater reliability was substantial for concordant symptoms provoked with hip flexion ($\kappa = 0.66$) and moderate for pain reproduced by the patellar compression test ($\kappa = 0.44$).

Palpatory test

Palpatory tests ranged in inter-rater reliability from moderate to poor (Table 4.4).

Physical performance measures

All of the physical performance measures showed good or substantial reliability for time and/or count. Pain report (numerical pain rating scale - NPRS), and rate of perceived exertion (RPE), show good or moderate reliability. Only the forty metre walk test met assumptions for a Bland Altman plot (Figure 4.8) which shows that differences between raters are highly likely to be less than five seconds.

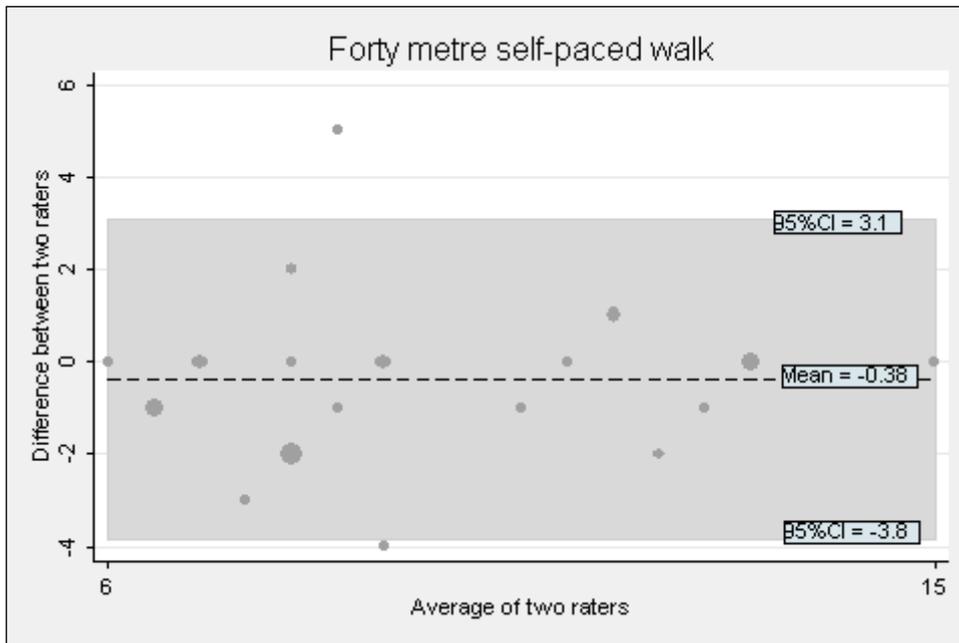


Figure 4.8 Bland Altman plot showing agreement between two raters for forty metre self-paced walk (in sec)

4.4. Discussion

Results from a review of the literature, and from the current investigation indicated numerous variables with at least moderate inter-rater reliability, which means they can be used with confidence to identify baseline predictors of outcome in a knee OA population. Inter-rater reliability of clinical measures exhibits a wider range of inter-rater agreement than measures examining patient history and symptoms, suggesting care should be taken when selecting clinical measures to include in a standardised baseline assessment.

Some of the results from the current investigation are similar to findings in previous studies; for example, hip joint range of motion (ROM) measured by goniometry was found to have good inter-rater reliability in the current study, similar to good inter-reliability reported for knee ROM in previous studies (Watkins et al., 1991; Fritz et al., 1998; Brosseau et al., 2001). Results from the current investigation are also consistent with the literature for the patella

compression test, which has moderate inter-rater reliability ($\kappa = 0.44$) comparable with published evidence ($\kappa = 0.43$) (Wood et al., 2006).

Accumulated evidence from a number of studies increases confidence in the findings and helps with selection of the most appropriate tests for the baseline clinical assessment for the prognostic study.

Variables that differed in the current study compared with published evidence included muscle strength using hand-held dynamometers (HHD). Knee flexors and extensors recorded moderate reliability (ICC 0.68 – 0.72), compared with good reliability in an elderly population with low back pain (ICC 0.87 – 0.93) (Dunn and Iversen, 2003). These results suggest that existing knee pathology or joint pain may impact the degree of reliability. Other differences included moderate ($\kappa = 0.6$) reliability of joint line tenderness, compared with fair ($\kappa = 0.21$ – 0.35) in other studies (Jones et al., 1992; Dervin et al., 2001; Wood et al., 2006). A stronger result in the current study may have been due to standardization of the procedure and a period of familiarization with test performance that was conducted prior to reliability testing. It is important for tenderness on palpation of joint lines to be reliable as it was one of the inclusion criteria for the clinical trial (Altman et al., 1986; Abbott et al., 2009).

4.4.1. Limitations and implications for future research

One of the limitations of the reliability investigation, and a factor that needs to be considered when reviewing the literature, is the recognized paradox of reported low kappa values with high percent agreement when low true prevalence exists in the target population (Feinstein and Cicchetti, 1990; Tooth and Ottenbacher, 2004; Sim and Wright, 2005). This is exemplified by the wipe test for knee swelling. Very few patients in this knee OA population demonstrated any swelling in the knee and therefore the wipe test was always counted as negative by the two raters, resulting in a low kappa value but high

percent agreement ($\kappa = -0.05$, % agreement = 89.67). To improve the interpretation of results, therefore, it is essential to also report percent agreement whenever the kappa statistic is used (Tooth and Ottenbacher, 2004).

A related problem is the study sample included patients with hip OA or knee OA. All tests were performed regardless of the joint affected which will have increased the number of negative findings i.e. hip subjects are likely to test negative to knee specific tests and vice versa. As stated above this will have lowered the prevalence of some clinical features, and may have influenced kappa values. It may be beneficial for future studies to perform reliability analysis for clinical tests pertinent to hip and knee OA on separate samples of participants with the relevant disease.

Skewed distribution of data is another very likely cause of the often-seen paradox of high percent agreement but low kappa as demonstrated in the test for crepitus ($\kappa = 0.22$, % agreement = 68.97) (Feinstein and Cicchetti, 1990; Lantz, 1997). Kappa can only approach 1.0 if both raters have exactly the same number of cases in each category or level, which is rarely the case in a clinical sample (Streiner, 1995). In fact, the maximum value of kappa can be as low as 0.1 or 0.2 in a sample in which distribution of these categories is heavily skewed (Streiner, 1995). In clinical situations in which skewing is present only to a small degree, a kappa value of 0.8 could be considered almost perfect agreement (Streiner, 1995). This has a large impact on the relevance of the standard interpretations of kappa values, and a very large impact on the validity and interpretation of much of the literature available on the reliability of physical examination items (Feinstein and Cicchetti, 1990).

There are a number of other limitations, some of which were present in the reviewed studies as well as the current study. Lack of adequate standardization in the methods and interpretation of test results may result in poor inter-rater

reliability. Technical difficulties included the difficulty providing adequate fixation of the lower limb during muscle testing of hip or knee extensors. The solution was to enrol the assistance of another examiner when testing potentially strong muscle groups. An unforeseen problem of inducing muscle cramps during testing of hamstring strength when the knee was at ninety degrees of flexion further affected some results. Consequently testing of hamstring strength should take place with the knee close to full extension, as muscle cramping was less problematic in this position.

Scheduling difficulties limited the ability to ensure that testing occurred at the same time of day for each patient. This potentially introduced measurement error, due to unknown variability in symptoms with the time of day. *A priori* it was decided testing would occur on two separate occasions, so as not to burden the patients with an additional hour of testing following the two hours of testing required for the clinical trial. In addition, it was thought a cumulative effect of testing may have led to a change in performance, due to fatigue or symptom aggravation, if testing was performed on the same day.

Tests were performed in the same order by both assessors. The assessment was designed to limit change in patient position, separate PPM to limit fatigue, and also to keep potentially aggravating tests (step-ups) until the end of the session. This over-rode any need to randomise order of testing for the second assessor. Patients were exposed to lots of tests, reducing likelihood of recall of responses.

The number of subjects tested was twenty-nine, four short of the sample size calculation. This potentially reduces the power of the study to produce statistically significant results. However, a review of the data confirmed that subjects demonstrated a good degree of heterogeneity in terms of age, severity of symptoms, and site of condition, suggesting the sample contained a good cross-

section of the population, meaning results are more generalisable. Also the pre-selected levels of agreement ($p_0=0.40$, $p_1=0.70$) were robust and therefore allow confidence in the interpretation of results.

Clinical measures had a greater range of inter-rater reliability results than variables obtained from patient history or self-reported symptoms. Additional efforts such as training of examiners or provision of operational instructions could possibly improve inter-rater reliability, and could be evaluated in future research.

4.5. Conclusion

The findings of the investigation into the inter-rater reliability of clinical measures supplement the information gained from reviewing the literature. Many variables have at least moderate inter-rater reliability, an important finding for their use as predictors of outcome. Inter-rater reliability will be used as a tool to screen variables for inclusion in the baseline assessment of the prognostic study.

5. Study to identify predictors of outcome: Methods

5.1. Chapter overview

The literature review (Chapter 2) identified patient characteristics that predict progression of knee osteoarthritis (OA). However, few studies have investigated progression or outcome of knee OA patients following physiotherapy intervention. The current study was designed to address this gap. Specifically the study was designed to investigate the hypotheses that baseline variables can be identified as predictors of successful outcome at one year following physiotherapy intervention for individuals with knee OA; and that probability of success at one year can be predicted by presence or absence of clusters of predictors (models).

Important components of the investigation were construction of a baseline assessment incorporating potential predictive variables; establishing the reliability of those variables; provision of standardised physiotherapy intervention; and assessment of outcome. The rationale for inclusion of variables in the baseline assessment is outlined in the following pages. The previous chapter (Chapter 4) has investigated inter-rater reliability of clinical measures and results will be incorporated here. Standardised physiotherapy intervention is summarised. Outcome was assessed at one year. Selected outcome measures and associated response criteria are described.

Data analyses were performed to:

- Identify a set of variables that would predict the likelihood of successful outcome at one year following physiotherapy intervention.
- Calculate change in post-test probability using the optimal model of baseline predictors.

- Investigate if specific combinations of named predictors produced different probabilities of success.

Three secondary analyses were conducted.

- The first investigated whether variables predicted treatment response or change due to natural course of disease over a one year period.
- The second examined nine week treatment response as a predictor of outcome at one year.
- The third aimed to identify predictors of poor outcome following physiotherapy intervention.

5.2. Study design

This was a prospective cohort study. Participants were drawn from the treatment arms of the larger Management of Osteoarthritis (MOA) randomised controlled trial (RCT) investigating the effectiveness of physiotherapy for the management of hip and knee OA (Abbott et al., 2009). Ethical approval was granted by the Lower South Regional Ethics Committee of the New Zealand Ministry of Health (ethics reference: LRS/07/11/044). (Appendix I). At the initial visit, potential participants gave informed consent, prior to being screened for eligibility. Examiners conducted a comprehensive assessment of potential predictor variables, and evaluated cardiovascular and other risks of participation. Details of the methods for the MOA clinical trial are available elsewhere (Abbott et al., 2009). Salient points are that randomisation occurred after the initial assessment, ensuring that examiners were blinded to group allocation; participants received standardised physiotherapy intervention; follow-up visits were conducted by the same blinded examiners at nine weeks, twenty-six weeks and one year, with the primary outcome assessed at the final visit. As OA is a chronic progressive condition, outcome assessment needed to be sufficiently long-term to identify relevant change. One year was considered to be adequate

for this purpose and matched the period selected for follow-up in similar trials (Deyle et al., 2000, 2005).

Sample size was calculated for the MOA trial. The aim was to recruit 224 subjects with either hip or knee OA: with a prudent allowance for dropouts of 20%, this would provide 95% power to detect differences in main effects (Abbott et al., 2009). One hundred and thirteen participants with knee OA were recruited to the MOA trial. Three out of four participants were randomised to treatment groups resulting in a cohort of eighty-five participants for the predictive study. The remainder ($n = 28$) formed the usual care group, who did not receive any intervention. There is no generally accepted method for calculating sample size for prognostic studies; however, design of the baseline assessment has to balance the inclusion of all potential predictors of outcome with the increase in sample size required to prevent overfitting of the model (Childs and Cleland, 2006). A ratio of ten participants per predictive variable in the final model is recommended to counter the risk of overfitting (Beattie and Nelson, 2006; Childs and Cleland, 2006).

5.2.1. Study setting

The MOA trial recruited adults with hip and knee OA from the Orthopaedic Department at Dunedin Public Hospital, and from general practices in Dunedin and surrounding areas. Dunedin, New Zealand has a population of 118,683 people, mostly in an urban setting with a rural hinterland (Statistics New Zealand, 2006). Recruitment took place from March 2008 – March 2009, with one year follow-up assessments continuing until March 2010. All assessments, follow-up visits and treatment sessions were conducted at the School of Physiotherapy, University of Otago, Dunedin, New Zealand.

5.2.2. Participants

All patients with a medical diagnosis of hip or knee OA referred to the Orthopaedic Department, or identified from General Practice databases were considered as potential participants in the MOA trial. Information outlining the trial was provided in the clinics. A research nurse performed an initial telephone interview to confirm eligibility, explain the trial, and ascertain ability to comply with trial commitments. Patients who passed this initial screen were then invited to attend for assessment at the School of Physiotherapy.

At the baseline assessment, details of the trial, including time required for attendance at treatment sessions and follow-up assessments, were reiterated before informed consent was obtained. Eligibility was confirmed using inclusion and exclusion criteria (Table 5.1).

At each stage of the recruitment process numbers of participants approached, screened, and eligible were documented to determine response rates.

Table 5.1 Inclusion/exclusion criteria for the MOA clinical trial

Inclusion Criteria: ACR clinical criteria for classification of Knee OA (Altman et al., 1986)	Exclusion Criteria
Knee Pain Plus ≥ 3 of following: <ul style="list-style-type: none"> • Age > 50 years • Crepitus • No palpable warmth • Tenderness on joint palpation • Bony enlargement • Morning stiffness < 30 minutes 	<ul style="list-style-type: none"> • Neurogenic disorder • Previous joint surgery to knee (excludes arthroscopy) • Rheumatoid arthritis • Knee joint injection prior 3 months • Current oral steroid use • Recent initiation opioid analgesia/arthritic drug (< 30 days) • Uncontrolled hypertension • Severe osteoporosis • Lower limb surgery < 6 months • Poor vision • Debilitating low back pain • Weight > 155 kg • Unable to walk > 10m unaided • Stated inability to complete proposed course of intervention and follow-up • Insufficient language skills to complete assessment tools • Insufficient comprehension to comply with interventions

Inclusion Criteria

The American College of Rheumatology (ACR) clinical criteria for classification of knee OA (Altman et al., 1986) were used as the basis for inclusion criteria (Table 5.1). The ACR criteria have been used extensively in previous research. Currently there is no single accepted method or “reference standard” for diagnosis of knee OA, although clinical signs and symptoms can be used to

make a confident diagnosis (Zhang et al., 2010). Radiographs are a useful adjunct but not necessary to make a diagnosis of knee OA (Zhang et al., 2010) (Bedson and Croft, 2008). The relationship between clinical symptoms and radiographic change has been disputed, with a recent systematic review reporting variation in the extent of correlation or discordance reported in different studies (Bedson and Croft, 2008). It is suggested that discordance may be greater in milder cases of the disease, with radiographic changes being more common in those with severe symptoms (Duncan et al., 2007; Bedson and Croft, 2008). The use of clinical criteria in the current study increased the likelihood of including participants in the early stage of disease before structural changes occurred, as well as those with a well-established diagnosis, resulting in a sample representing the full spectrum of disease. The clinical criteria used here are also easily assessed in usual clinical practice by doctors or physiotherapists, which is an additional benefit in terms of external validity.

Participants nominated one knee to be the primary knee for investigation. This is referred to as the index knee, and in the majority of cases the index knee produced the worst symptoms.

Exclusion Criteria

Exclusion criteria are shown in Table 5.1. The intention was to recruit as broad a cross-section of the community as possible and not to exclude people unnecessarily. This would better reflect a typical community of patients and make the results more generalisable. Items were selected to ensure potential participants did not have underlying medical conditions, or were not undergoing treatment that could influence response to treatment or bias outcomes. Other items addressed included: possible contraindications to the treatment programmes; factors that could affect the capacity of participants to comply with treatment; and factors affecting patient safety.

5.2.3. Assessors

The four assessors involved in the trial were all registered physiotherapists with a range of postgraduate training and experience. This is reflective of a typical physiotherapy clinical environment. All assessors underwent a period of training and supervised performance of procedures to achieve familiarity with assessment techniques and to increase consistency. The author was the primary assessor for the participants with knee OA. Assessors were blinded to group allocation and blinding was maintained for follow-up assessment.

5.3. Baseline assessment

The baseline assessment covered a range of variables, including elements from patient history and demographic information, as well as physical measures (Appendix H). It is recommended that an exhaustive list of potential predictors should be generated from previous relevant research as well as expert opinion or consensus (Beattie and Nelson, 2006). However, it may not be possible to include all potential predictors, as the increased sample size required to prevent overfitting the model may not be practicable, or affordable (Childs and Cleland, 2006; Moons et al., 2009).

The baseline assessment consisted of two self-report patient questionnaires and an interview/examination conducted by an assessor. The assessor was available to help with the completion of the questionnaires, to answer queries from the participants, and to check all questions were answered. For the physical examination, the underlying philosophy was to select measures that were easy to perform, in common clinical usage, and required only basic equipment. This would make findings more easily transferrable to the clinical environment. The number of baseline variables had to be restricted to keep assessment within a reasonable timeframe for participants, and to limit aggravation of symptoms or fatigue by over-testing.

The physical examination was constructed with reference to published evidence about variables and their association with development or progression of knee OA; typical clinical features of knee OA; or features of knee OA with potential for modification by physiotherapy intervention. Predictors of knee OA progression identified in the systematic review were considered for inclusion. Information regarding the inter-rater reliability of clinical measures was also considered (Chapter 4). When there were a range of techniques or measures available for testing a single clinical feature, the most reliable technique was selected. Similar consideration was made in selecting variables from the questionnaires about patient demographics, clinical history and disease characteristics, demonstrating a sound theoretical rationale for including variables as potential predictors, and limiting the possibility of spurious findings (Kraemer et al., 2002; Stanton et al., 2010).

A manual of operating procedures was produced describing tests and standardising their execution (Appendix K). In addition to instructions for performing tests, it included descriptions of equipment; definitions of terms; criteria for responses to tests; patient positioning; and any standardised instructions to the patient e.g. for producing muscle contraction.

It should be noted that forms included many variables that were required for the MOA clinical trial or a related study on hip OA, but were not necessary for this current prognostic study of knee OA. A full description of measures used in the MOA trial is provided elsewhere (Abbott et al., 2009). The following sections describe collection of data relevant to the identification of predictive variables for knee OA outcome following physiotherapy intervention.

5.3.1. Patient history

Patient history covers all the patient characteristics relating to demographic information, psychosocial factors, disease history, and symptoms.

Demographic information

This included age, sex, and ethnicity.

Psychosocial factors

Psychosocial factors are considered to be important and often neglected features affecting osteoarthritis and outcomes (Axford et al., 2008; Hunter and Lo, 2008; Zhang et al., 2008; Somers et al., 2009). They have been investigated as potential predictors of natural progression, although as shown by review of the literature (Chapter 2), the evidence is limited for their role (Sharma et al., 2003a). Consequently, assessment of self-efficacy, fear avoidance, catastrophizing, and depression were included at baseline assessment as possible predictors of outcome in this study.

The instruments used were the Pain Belief Screening Instrument (PBSI) for evaluation of self-efficacy, fear avoidance, and catastrophizing (Sandborgh et al., 2007), and the two-item case-finding instrument for detection of depression (Whooley et al., 1997). Both instruments have good psychometric properties, with the latter being recommended for use in patients with chronic physical health problems (Whooley et al., 1997; Sandborgh et al., 2007; Pilling et al., 2009).

Other psychosocial factors included marital status, level of education, income, work status, smoking history, and self-reported activity level.

Moderate physical activity does not predict progression of knee OA, may have a protective effect, and is likely to result in decreased pain and increased function (Schouten et al., 1992; Cooper et al., 2000; Sharma et al., 2003a; Bosomworth, 2009). However, it is recognised that pain and physical limitations of knee OA, in addition to other barriers, may restrict participation in physical activity and thereby adversely affect functional ability (Bennell et al., 2004; Rosemann et al., 2008). Physical activity is not often measured in clinical trials, partly due to the difficulty in obtaining accurate information (Bennell et al., 2004). While there are a number of physical activity questionnaires available (Hussey

and Wilson, 2003), most require the completion of a set of questions estimated to take approximately five minutes for older participants to complete (Washburn et al., 1993; Bennell et al., 2004). This was considered to be too time-consuming for the current study. In the absence of a suitable one or two-item validated tool, a single question was formulated to assess physical activity (Appendix H, p.2).

Disease history/history of injury

Information relating to disease history was specific to the index knee, unless otherwise stated. It included: symptom duration; previous knee injury; previous lower limb injury; history of falls; and previous helpful physiotherapy treatment. Shorter symptom duration (< 1 year) has been identified as a predictor of successful response to physiotherapy treatment in patients with hip OA (Wright et al., 2011). No evidence was found that associated duration of symptoms with treatment response for patients with knee OA. However, longer duration of symptoms has been associated with poorer clinical outcome in older patients with knee pain (Mallen et al., 2007b). Symptom duration was therefore considered important in the current study. Despite there being strong evidence that previous knee injury does not predict progression of knee OA, it is an established risk factor for incident knee OA, suggesting different processes are operating at different stages of the disease (Cooper et al., 2000). Its importance in influencing treatment response is unclear and was therefore investigated.

Symptoms

One of the main symptoms of knee OA is pain, causing many (though not all) patients to consult with primary health care practitioners (Bedson et al., 2007). Alleviation of pain is frequently a target for intervention, including physiotherapy. Baseline pain was therefore considered to have potential as a predictor of outcome following physiotherapy intervention, even though evidence from the systematic review about the influence of pain on natural progression of knee OA was conflicting (Chapter 2). A number of questions in the

baseline assessment examined different aspects of pain presentation. The numerical pain rating scale (NPRS) has been used reliably in an elderly population with musculoskeletal pain (Mawdsley et al., 2002).

Instability has been recognised as a problem in knee OA, and may be reported as a sensation of “shifting, buckling or giving way of the knee” (Irrgang et al., 1998; Fitzgerald et al., 2004a). Its influence on daily functioning can be assessed using a question from the Knee Outcome Survey-Activities of Daily Living Scale (KOS-ADLS)(Irrgang et al., 1998). This self-reported measure has been shown to be reliable in patients with knee OA (Fitzgerald et al., 2004a). Studies have also identified instability as a predictor of function for patients with medial knee OA that may be accommodated for by activity modification (Schmitt and Rudolph, 2008). As instability is potentially modifiable it may also be amenable to treatment (Fitzgerald et al., 2004a). It was therefore investigated as a possible baseline predictor of outcome following physiotherapy intervention.

There is some evidence, both published and empirical, suggesting irritability of a condition influences patient outcomes, and is frequently used by physiotherapists when making treatment decisions (Hurley et al., 2002; Smart and Doody, 2007) (Chapter 3). It was therefore decided to formally assess irritability of the index knee, partly to ensure the safe application of treatment interventions, but also as a possible predictor of outcome. The i-score was developed as a standardised procedure for assessing and reporting irritability (Chapter 3).

Simple standardised questions addressed other potentially important symptoms including: stiffness; locking; heat; and swelling (Appendix H, pp.4-5).

5.3.2. Clinical features and physical examination

Items included in the physical examination were selected from orthopaedic text books describing standard knee examinations (Magee, 2002; Cleland, 2005; Malanga and Nadler, 2006), in addition to tests from studies of composite

assessment of knee OA (Cushnaghan et al., 1990; Jones et al., 1992; Cibere et al., 2004; Wood et al., 2006); these are described in the Manual of Operating Procedures (MOP) (Appendix K). There are often a number of tests available to examine the same clinical feature. When more than one test existed, selection was based on documented use in a similar population; tests with the highest inter-rater reliability; and tests that are easily employed in a typical clinical setting. When there was no report of reliability for a specific test, inter-rater reliability was investigated in a separate study; results from this, and all previously established reliability values are reported in Chapter 4.

The physical examination was designed to limit the number of times participants had to change position. Consideration was given to patient fatigue, allowing recovery time from more strenuous activity e.g. Forty-metre self-paced walk (SPW), and restricting the total time for the physical examination to less than one hour. Also tests more likely to aggravate patient symptoms (e.g. step-ups) were left to the end of the examination to prevent biasing other responses. The order of testing was the same for all participants.

Obesity

The evidence that obesity predicts progression of knee OA is conflicting, although it appears more likely to be a factor in the long-term (Chapter 2). However, there is high quality evidence that weight reduction can produce good outcomes for patients with knee OA (Jamtvedt et al., 2008a). There are different methods of evaluating body composition and fat distribution, although a recent study concluded that none of these alternate measures offered any advantage over the simple calculation of body mass index (BMI) or weight, when investigating risk factors for radiographic knee OA (Abbate et al., 2006). BMI has the additional benefit of being widely used in research and therefore enables comparison of results. For these reasons BMI was selected as the preferred measure of obesity in this study.

Alignment and position

A detailed review identified strong evidence that malalignment of the tibio-femoral joint in the frontal plane (valgus/varus deformity) is a predictor of knee OA progression (Tanamas et al., 2009). Additionally it can affect forces across the knee joint, and contribute to individual variations in treatment response to interventions such as exercise (Gibson et al., 2010). The gold standard for assessment of alignment is considered to be the standing long leg radiograph, although this method can be time consuming, expensive, and lacks clinical utility for primary health care practitioners (Hinman et al., 2006; Brouwer et al., 2007). Numerous clinical measures have been described (Cibere et al., 2004; Hinman et al., 2006). Visual inspection was selected as it is quick and easy to perform, requiring no specialist equipment or measurement procedures (Magee, 2002; Malanga and Nadler, 2006). While inter-rater reliability was found to be “fair” in the previous study (Chapter 4), the recognised importance of knee alignment to progression of knee OA and treatment outcome justified inclusion in the baseline assessment.

Excessive pronation or supination of the foot, especially during mid-stance of the gait cycle, influences force transmission and loading of the knee, thus contributing to OA (McPoil and Cornwall, 2005; Reilly et al., 2009). Measurement of the longitudinal arch angle (LAA) in a static position can predict dynamic posture at mid-stance, and is easily performed using a short-arm goniometer and bony landmarks (Dahle et al., 1991; McPoil and Cornwall, 2005).

Patella malalignment has been associated with progression of patello-femoral OA in a radiographic study (Hunter et al., 2007). No studies were found that used clinical tests of patella malalignment to predict OA progression. The method selected for this study was adapted from McKewan et al. (2007) and assessed medio-lateral displacement only.

Joint movement

Range of motion (ROM) is commonly assessed by physiotherapists. Limitation of joint movement is a common feature of knee OA, specifically flexion and extension. It is linked with pain and functional impairment and is a logical therapeutic target for physiotherapy interventions.

Measures of ROM were performed on the knee, hip, and ankle, and are described in the MOP (Appendix K). All movements were performed actively by the patient and then assisted to end of available range by the assessor. Concordant pain was recorded. Additional information was gathered about end-feel and pain resistance sequence, as they are purported to reflect the acuity of a condition and guide treatment selection (Cyriax, 1982; Maitland, 1986). Research has questioned the validity and inter-rater reliability of these concepts (Hayes et al., 1994; Fritz et al., 1998), although it is conceivable that treatment response may be mediated by baseline end feel and pain resistance sequence.

Length of muscle and non-contractile tissues was examined using the 90/90 test for hamstring length, and the Thomas test for hip flexor contracture and rectus femoris tightness (Bandy and Irion, 1994; Magee, 2002). Stretching has been described as a useful target for physiotherapy intervention for knee OA (Reid and McNair, 2010) suggesting baseline soft tissue restriction could potentially be predictive of good outcome.

Accessory movements

Physiotherapists use accessory movements to identify and treat painful or tight structures (Maitland, 1991). If pain and/or resistance are detected in OA knee, they are potentially modifiable. The extent to which this baseline restriction will predict outcome is unknown.

Muscle strength

Evidence concerning muscle strength as a predictor of natural progression of knee OA is conflicting (Chapter 2), and appears to be complicated by the

confounding effects of alignment, joint laxity, and instability (Sharma et al., 2003b; van der Esch et al., 2006; Schmitt and Rudolph, 2008). This complex interaction also affects response to a strengthening programme (Lim et al., 2008). However, exercise, including muscle strengthening, is consistently recommended as an essential component of conservative management (Ottawa Panel, 2005; National Collaborating Centre for Chronic Conditions, 2008; Zhang et al., 2008). One study showed that quadriceps and hamstring strength training reduced radiographic progression of knee OA over thirty months (Mikesky et al., 2006). Quadriceps and hamstring strength were therefore evaluated as potential predictors of outcome in this study.

Hip muscle strength was assessed as it has the potential to affect loading of the knee joint (Bennell et al., 2007). A recent study demonstrated significant difference in strength of all hip muscle groups between subjects with knee OA and asymptomatic individuals, with the greatest difference being for hip external rotators (Hinman et al., 2010). Weak hip abductors can increase progression of knee OA (Chang et al., 2005), and knee OA patients have exhibited greater hip adductor strength (Yamada et al., 2001). The role of other hip muscles in progression of knee OA, or predicting response to treatment, is unclear.

Ankle plantarflexion strength was assessed because of the dual role of the gastrocnemius muscle in plantarflexion of the ankle and extension of the knee. The straight leg raise (SLR), where the patient is asked to lift the whole lower limb with the knee in full extension, was included as a quick test of quadriceps muscle strength (Magee, 2002).

Hand held dynamometry (HHD) was selected as the method for assessing muscle strength. Hand held dynamometers are inexpensive, portable, easy to use, and provide an objective measure of muscle strength (Ford-Smith et al., 2001; Dunn and Iversen, 2003). HHD performs well when correlated with manual

muscle testing (Bohannon, 2001). Additionally, HHD has been used reliably with healthy older individuals, and with elderly patients with orthopaedic problems (Ford-Smith et al., 2001; Dunn and Iversen, 2003; Roy and Doherty, 2004). The HHD used in this study was the Nicholas MMT, (Model 01160, Lafayette Instruments, Lafayette, Indiana).

HHD as used in the current study provided measures of muscle strength in kilograms (kg). This was normalised for body weight or mass (m), incorporating an allometric parameter of 2/3 (0.67) to account for the fact that muscle strength increases at a lower rate than body mass (Jaric, 2002). The equation used for normalisation was:

$$S_n = S/m^b$$

where S_n = normalised strength, S = recorded muscle force, m = body mass and b = allometric parameter = 0.67 (Jaric, 2002).

Ligament and meniscus tests

Valgus/varus laxity is associated with knee OA (van der Esch et al., 2005) with some evidence that it can predict progression (Dieppe et al., 1993; Sharma et al., 2003a). Meniscal damage was identified in 75% of patients with knee OA with no history of trauma (Berthiaume et al., 2005) and has been shown to predict future cartilage loss quantified by magnetic resonance imaging (MRI) analysis (Sharma et al., 2008). Clinical measures of these features were included in baseline assessment (Appendix H, p.11).

Palpation tests

The knee joint was palpated for joint line tenderness, bony enlargement, heat and crepitus, to identify inclusion criteria (Altman et al., 1986). There is some evidence for the role of crepitus, heat, and effusion as predictors of natural progression (Dieppe et al., 1993; Ledingham et al., 1995), but no evidence that it influences outcome following physiotherapy intervention.

Patello-femoral OA contributes to the symptoms of generalised knee OA in addition to being a separate clinical entity (Hinman and Crossley, 2007; Duncan et al., 2009). Its role in disease progression has not been determined. Baseline assessment included a pain provocation compression test, palpation for joint line tenderness, as well as positional assessment (Fitzgerald and McClure, 1995; Cleland, 2005).

Physical Performance Measures (PPM)

Physical performance measures (PPM) are standardised tests of functional activities, evaluated by predetermined criteria such as time or number of repetitions (Stratford et al., 2003). They provide useful information about physical limitations associated with OA, and how these contribute to impairment and disability (Lin et al., 2001). Additionally they can be used to evaluate change and predict outcomes (Jette et al., 1999). A battery of tests is recommended, as patients will experience different problems according to the stage or severity of their disease, and to ensure thorough examination of all underlying constructs e.g. physical functioning, ROM, strength, pain (Stratford et al., 2003; Piva et al., 2004; Terwee et al., 2006). Correlation of PPMs and self-report measures may be enhanced by including patient reports of perceived exertion and pain during test performance (Stratford et al., 2003). These recommendations, in addition to established validity and reliability, influenced the selection of PPMs for this study. Included PPMs were: forty-metre self-paced walk (SPW); timed up and go (TUG); thirty-second sit to stand; twenty-centimetre step test.

The forty-metre SPW was adapted to suit the research facility and consisted of four ten-metre lengths with three turns, rather than the previously documented two twenty-metre lengths with one turn (Kennedy et al., 2005). The TUG was performed as described by the originator of the test (Podsiadlo and Richardson, 1991) and used since in OA studies (Stratford et al., 2006). The thirty-second sit to stand was developed as a response to the floor effect of the chair stand test, in

which some of the elderly population were unable to complete the test (Jones et al., 1992; Guralnik et al., 1994). The twenty-centimetre step test is one of the rising and descending tasks included in the DynaPort knee test, and has good validity and reliability in a knee OA population (van den Dikkenberg et al., 2002; Mokkink et al., 2005).

5.4. Overview of trial procedure

5.4.1. Randomisation

Participants who met the inclusion criteria for the MOA trial were randomised following baseline assessment, according to the trial protocol (Abbott et al., 2009). An online method of randomisation was used by the trial administrator, which ensured assessors remained blinded to group allocation.

5.4.2. Standardised treatment

Participants were allocated to one of four groups: usual care; manual therapy; exercise therapy; or both exercise and manual therapy. Exercise therapy included a ten-minute warm-up (usually on a stationary bicycle), stretching exercises, strengthening exercises and neuromuscular exercises incorporating balance and coordination activities. Manual therapy included accessory movements of the joints, physiological movements and manual stretches applied by the physiotherapist, as well as massage. Relevant points are that the interventions were standardised with knee participants receiving a package of interventions according to group allocation, in addition to secondary interventions based on impairment criteria (Appendix N). Any deviations from the treatment protocol were documented and discussed. All treatment sessions were performed by physiotherapists at the School of Physiotherapy, University of Otago, Dunedin. Assessors did not provide any treatment interventions, to ensure maintenance of blinding. Treatment guidelines and manuals were

provided and the treating physiotherapists received training prior to commencement of treatment sessions. Treatment sessions were limited to one-hour contact time with the physiotherapist, once per week, for a period of seven weeks, followed by two booster sessions at fifteen to sixteen weeks approximately. Participants in all treatment groups also had a home exercise programme which they were asked to perform daily, and to keep an exercise log book, so adherence could be assessed. At completion of the physiotherapy intervention, participants were encouraged to continue with their home exercise programmes, and submit log books regularly, to monitor adherence and provide motivation.

The usual care group (UC) did not receive any of the physiotherapy interventions. However, they could pursue any form of treatment independently, or as directed by their medical practitioner, including physiotherapy from another provider, hydrotherapy, exercise, medication, surgery and so forth. Details of any treatment undertaken during the study period were collected at follow-up assessments. All participants received newsletters with trial information.

The current study was not concerned with effectiveness of different forms of physiotherapy treatment; therefore all three treatment groups were combined to form the prognostic study cohort.

5.4.3. Follow-up assessments

Participants attended for follow-up assessments for the MOA trial at nine weeks, twenty-six weeks, and one year. Information relevant to the prognostic study was assessment of outcome at nine weeks and one year. The assessors remained blinded to group allocation and participants were coached not to divulge information that might un-blind the assessors. Assessors recorded all instances when participants inadvertently revealed their group allocation.

Follow-up assessments were adapted from the baseline assessment and consisted of questionnaires and a physical examination. The assessor ensured the nominated index knee from baseline remained the index knee at follow-up, regardless of which knee was more symptomatic on the day. Participants who had undergone joint replacement surgery during the study period were encouraged to remain in the trial and attend for follow-up assessments. Type and date of surgery were noted.

5.5. Outcome measures

Outcome at one year was dichotomized into success or non-success with physiotherapy intervention. The two outcome measures used were the Western Ontario and McMaster osteoarthritis index (WOMAC 3.1) and the global rating of change (GRC) (Appendix M).

The WOMAC has been used extensively in OA research and has demonstrated validity and responsiveness (Bellamy et al., 1988; Angst et al., 2001). It consists of three subscales evaluating pain, stiffness, and function, thus covering multiple domains, as recommended for evaluating clinical importance of outcome (Dworkin et al., 2008). The version used in the MOA trial used Likert scales of 0 to 10 for each question. Maximum score for the pain subscale was 50; for the function subscale 170; and for the stiffness subscale 20. Maximum possible score for the combined WOMAC index was 240.

The GRC is an anchor-based method that has been used previously to determine successful outcome following physiotherapy intervention (Leshner et al., 2006; Cleland et al., 2007; Currier et al., 2007). Anchor based measures are a method advocated for detecting minimal clinically important difference (MCID) (Dworkin et al., 2008; Terwee et al., 2010). These use an external criterion or standard to define change that is perceived as important from the perspective of the patient, or other stakeholders (Jaeschke et al., 1989). The version used in the

MOA trial had fifteen categories ranging from “a very great deal worse” to “a very great deal better”.

5.6. Response Criteria

Response criteria have been developed by the Standing Committee for Clinical Trials Response Criteria Initiative and Outcome Measures in Rheumatology, and Osteoarthritis Research Society International (OMERACT-OARSI). These responder criteria dichotomize an individual patient’s response to intervention (Pham et al., 2004). They cover the domains of pain, function, and patient-reported improvement, and include both absolute and relative change in scores (See Table 5.2).

Table 5.2 OMERACT-OARSI responder criteria to define outcome at one year for participants with knee osteoarthritis following physiotherapy intervention.

<p>≥ 50% change in pain or function</p> <p>AND absolute change ≥ 20</p>	<p>Change in at least 2 of the following 3 scores:</p> <ul style="list-style-type: none"> • Pain ≥ 20% AND absolute change ≥ 10 • Function ≥ 20% AND absolute change ≥ 10 • Patient’s global assessment ≥ 20% AND absolute change ≥ 10
---	--

These criteria were applied to WOMAC scores to determine success, or not-success following physiotherapy intervention. The three subscales of WOMAC required conversion to scores out of 100, prior to determining values for absolute change.

As the GRC was only assessed at follow-up and could not therefore be reported as an absolute or relative change from baseline, a clinically important

cut-point had to be determined. However, there is no consensus or evidence to decide on cut-points for anchor-based methods (Terwee et al., 2010). Response according to the OMERACT-OARSI responder criteria requires a $\geq 20\%$ change in patient global assessment. The fifteen-point GRC scale used in this study has seven steps that represent improvement. A 20% improvement would equate to change of greater than two-steps resulting in a cut-point of \geq “somewhat better” (Appendix M). This has been used in a previous study of physiotherapy intervention (Currier et al., 2007).

5.7. Data management and preparation

All patient information and data were recorded on standardised forms and used unique patient identifying numbers. Following assessment, forms were scanned so all information could be stored and processed electronically. Details on data flow and storage are provided in the trial protocol (Abbott et al., 2009). Data for the prediction study was subject to manual checking and auditing procedures.

5.7.1. Management of missing data

Attrition of participants in longitudinal clinical trials is inevitable and will result in missing data. The best approach to dealing with missing data is to limit the number of withdrawals or dropouts (Unnebrink and Windeler, 2001; Walton, 2009).

It is recognized that failure to include dropouts in analysis can produce biased or misleading results as well as losing potentially useful information (Gadbury et al., 2003; Fielding et al., 2008; Molnar et al., 2008). Therefore the standard for clinical effectiveness trials is to perform an intention to treat analysis (ITT), meaning all participants are analysed according to their randomized group, even if they fail to complete treatment or do not attend for the final data collection (Liu and Gould, 2002). The ITT principle preserves the balanced distribution of

participants to groups created by randomization (Unnebrink and Windeler, 2001). ITT analysis was used in the MOA trial (Abbott et al., 2009).

Options considered for dealing with missing data included last observation carried forward (LOCF) and multiple imputations (MI). LOCF uses the last available measurement or value for a subject to replace the missing value. This method, though simple to perform, has a number of underlying assumptions. The primary assumption that the available measured value does not change between observations has been identified as a major limitation, especially in chronic, progressive disease (Liu and Gould, 2002; Haukoos and Newgard, 2007; Molnar et al., 2008). Furthermore, this assumption may not hold in clinical trials, such as the MOA trial, where interventions are likely to produce change (Prakash et al., 2008). Additionally, reasons for dropout are varied and unpredictable, creating either underestimation or overestimation of effect, which refutes the argument that LOCF produces conservative estimates of effect (Haukoos and Newgard, 2007; Prakash et al., 2008).

A well-recognized bias of LOCF is that carrying the same observation forward decreases the variability of the data while the sample size remains the same, thus mathematically exaggerating the statistical significance of the results (Liu and Gould, 2002; Gadbury et al.; 2003, Prakash et al., 2008).

The unpredictable nature of change in outcomes in people with knee OA, and their unknown response to treatment, in addition to bias introduced by underestimating variability of the data, makes LOCF unfavourable as an option for dealing with missing data in this study.

MI is a more sophisticated method for replacing missing values, and involves the generation of a range of plausible values based on the observed data from “not-missing” participants (Newgard and Haukoos, 2007). Techniques such as MI are increasingly recommended for dealing with missing data as these allow

for remaining uncertainty following imputation and preserve study power (Newgard and Haukoos, 2007; Prakash et al., 2008).

MI was selected *a priori* as the method for handling missing data in the MOA trial (Abbott et al., 2009). Therefore MI data for knee participants with missing outcome data at nine weeks and one year was available and was the preferred approach for the prognostic study.

5.7.2. Management of incomplete data due to surgical intervention

Major joint surgery during the study period was identified as a potential confounder of results, as outcome for these participants was strongly influenced by surgery rather than physiotherapy intervention. One option was to exclude these participants from the analysis. However as previously stated, this would violate the ITT principle for analysis and result in loss of information. If the one year follow-up data were regarded as missing data then imputation methods could be considered. However, this is a subset of participants with a specific reason for “missing data”, i.e. surgery: they are likely to have more severe disease or symptoms than the rest of the study population, hence the need for surgery. This potentially violates the main assumption for MI, which is that data are missing at random (Newgard and Haukoos, 2007; Walton, 2009). Additionally, there are the previously stated limitations of using a simple imputation method such as LOCF.

An alternative is to consider these participants as non-responders to physiotherapy intervention at one year, using surgical intervention as the outcome measure. Total knee replacement (TKR) has been used in some prognostic studies to define outcome for knee OA progression (Dieppe et al., 1993; Cicuttini et al., 2004). However, joint replacement has been criticised as a clinical outcome for trials of structure-modifying drugs in hip and knee OA, as there are no universally accepted guidelines of when to perform joint surgery

(Altman et al., 2005). Notwithstanding these reservations, it seems clinically justifiable to regard surgical participants as non-responders to physiotherapy intervention.

5.7.3. Screening of baseline variables for analysis

There were a large number of variables assessed at baseline (Appendix H). This number was reduced for analysis to prevent over-fitting of models (Dawson and Trapp, 2001). Screening and exclusion of variables considered the following recommendations: a ratio of five subjects per potential predictor variable to perform regression (Dawson and Trapp, 2001); a ratio of ten subjects per predictor variable for final models (Concato et al., 1993; Childs and Cleland, 2006); variables can be clinically justified (Fitzgerald, 2010; Stanton et al., 2010); high-quality data (Concato 1993; Childs 2006). Baseline variables were also excluded from analysis for the following reasons: inclusion/exclusion criteria; assessment of patient safety; supplementary information for baseline variables e.g. joint end-feel, rate of perceived exertion (RPE) for PPMs; non-discriminatory across cohort e.g. SLR; redundant information with two variables measuring the same domain, e.g. fixed flexion deformity and knee extension ROM. Additionally as participants were also part of a larger sample for a clinical trial, some variables in the baseline assessment related to research questions for different, un-related studies.

Some variables were combined or composite scores calculated meaning the contributing variables could be excluded e.g. height and weight were combined into BMI; the i-score was calculated from four other variables; PBSI was calculated from seven items. Reasons for exclusion of variables from analysis will be reported in the results section.

5.7.4. Cut-points in data

Baseline variables included binary, categorical, and continuous data types. Not all categorical data were ordinal. In prognostic studies investigators frequently dichotomize or create cut-points in data. While this allows for ease of analysis and easier interpretation of results, it can be problematic, leading to loss of information with an associated reduction in statistical power (Altman and Royston, 2006). If cut-points are to be used, their selection should be guided by biological reasoning and not data driven (Altman et al., 1994).

These issues influenced the decision in this study to retain data whenever possible. However, the aim was also for information to make sense and be useful in a clinical environment; for these reasons individual variables were examined with respect to their clinical utility and any justification for creating a cut-point. Decisions about creating cut-points and their calculation were done prior to other analyses.

Creation of cut-points in continuous data

Age and BMI are two variables that are frequently categorised. In addition they are potential confounders in knee OA studies. For these reasons they were thoroughly investigated for association with outcome in both continuous and categorised formats.

PPMs are designed for ease of assessment and understanding. A cut-point has potential to enhance clinical understanding for these measures, or to provide a threshold when a PPM is used for screening purposes.

Recognised cut-points were used when available: e.g. BMI was categorised using the World Health Organisation's (WHO) classification, where <18.5 kg/m² is underweight; $18.5 - 24.99$ kg/m² is normal; $\geq 25.0 - 29.99$ kg/m² is overweight: and ≥ 30.0 kg/m² is obese (World Health Organisation, 2006). Alternatively, receiver operating characteristic (ROC) curve analysis was used to identify optimal cut-points (Portney and Watkins, 2000).

A ROC curve plots the sensitivity (or true-positive rate) to 1 – specificity (the false-positive rate) for different choices of cut-off. A test (or variable) with perfect discriminative ability (i.e. 100% sensitive and 100% specific) would give an area under the curve (AUC) of 1. A diagonal plot (AUC = 0.5) indicates that the test (or variable) is unable to discriminate outcome. Data for selected continuous variables were plotted as ROC curves. For each variable, the point closest to the top left-hand corner was selected as the cut-off value with the greatest ability to discriminate outcome (Kirkwood and Sterne, 2003).

Creation of cut-points in ordinal/categorical data

Similarly for ordinal variables or categorical variables with more than two categories, the preference was to retain all the information. Variables were examined to determine whether collapsing categories made more clinical sense than retaining the data in its original form. Clinically justifiable cut points were created using published evidence when available.

5.8. Data analysis

5.8.1. Overview of analysis

Outcome was calculated for all participants as described in Sections 5.5 and 5.6. All three treatment groups were combined to form the prognostic study cohort. The fourth (UC) group was analysed separately with results available for comparison.

The analysis took place in several stages. Primary analysis used univariate analysis to identify baseline variables that had a *statistical association* with successful outcome at one year. These variables were entered into a multivariate analysis using logistic regression with backward elimination to identify a set of variables that *predicted* successful outcome at one year. Different numbers of identified predictor variables were grouped and regarded as a “test”. Post-test

probability of success for different groups was then calculated to determine the optimal number of variables required to predict successful outcome.

Odds ratios (OR) are reported for potential predictor variables, representing the odds of a successful outcome (responder) over the odds of not having a successful outcome (non-responder), in participants with knee OA. ORs are frequently reported in prognostic studies and will allow comparison with previous work on predictors of knee OA progression (see Chapter 2).

Likelihood ratios (LR) incorporate sensitivity and specificity information and are used to evaluate shifts in probability of an outcome following the application of a clinical test (Furukawa et al., 2008). A positive LR (LR+) indicates that given a positive test, there is an increase in probability of a successful outcome. A negative LR (LR-) indicates that given a negative test there is a decrease in probability of a successful outcome. The significance of different magnitudes of LRs are shown in Table 5.3 (Jaeschke et al., 1994).

Table 5.3 Magnitude of likelihood ratios (LR) and effect on post-test probability (Jaeschke et al., 1994)

Magnitude LR+	Magnitude LR-	Change in pre-test/ post-test probability
LR+ > 10	LR- < 0.1	Large and often conclusive change
LR+ 5 – 10	LR- 0.1 – 0.2	Moderate change
LR+ 2 – 5	LR- 0.2 – 0.5	Small but sometimes important change
LR+ 1 – 2	LR- 0.5 - 1	Small but rarely important change
LR+ = likelihood ratio following a positive test; LR - = likelihood ratio following a negative test		

In the current study a test consists of different numbers of predictors from the multivariate model. LRs were used to calculate change in probability of success when different numbers of predictor variables were present.

All statistical analyses were performed using STATA 10.1 (StataCorp, 4905 Lakeway Drive, College Station, Texas 77845, USA).

5.8.2. Univariate analysis

All potential predictor variables were analysed with simple logistic regression, as there was a binary outcome. Variables were retained for multivariate analysis if $p < 0.2$. This generous level of significance avoided excluding potentially useful variables too early in the analysis and equates to similar studies (Jellema et al., 2006; Weigl et al., 2006; Vincenzino et al., 2010).

Univariate analysis was performed using baseline variables as the independent variables. The dependent variable for the primary analysis was successful outcome at one year.

5.8.3. Multivariate analysis

Retained variables were entered into a multivariate logistic regression, with backwards elimination. Variables with $p > 0.1$ were eligible for removal from the model at each step; the variable with the highest p-level was manually removed from the model until the model ceased to improve. This was detected by an increase in the p-value for the whole model. Additionally Akaike information criterion (AIC) and Bayesian information criterion (BIC) statistics were monitored. These are measures that combine fit and complexity of models. An increase of either AIC or BIC indicated a weaker model, resulting in selection of the previous iteration as the final model.

Post-estimation tests

Post-estimation tests included the Hosmer Lemeshow goodness of fit (GOF) test in which a chi-squared test is used to compare whether actual (observed)

values are significantly different to those predicted by the model (expected). If the p-value is large there is no significant difference, indicating a good fit of the model to the data. Collinearity was also tested, as inclusion in the final model of two highly correlated variables ($r > 0.5$) is problematic. It can lead to an erroneous conclusion that neither variable has a strong association with outcome when individually they may each have a strong association.

ROC curves were generated for all models. AUC was calculated and provides a measure describing the ability of a model to discriminate between successful and not-successful outcome at one year. An AUC value of 1.0 describes perfect discrimination, while 0.5 indicates the model does not discriminate beyond chance. Additionally the ROC curve plots a range of probability values and allows a cut-off value of probability to be determined – usually the point closest to the top left-hand corner of the chart. Alternatively associated tables can be consulted to determine the cut-off based on sensitivity or specificity values, LR's or percentage of accurate classification.

5.8.4. Confounders

Age, BMI, and sex are all potential confounders in knee OA studies. Each was examined for a univariate association with outcome. Any confounders retained after multivariate analysis were used to perform further stratified multivariate analysis.

5.8.5. Combinations of predictor variables and post-test probability

Pre-test probability was calculated as the percentage of participants who responded to physiotherapy intervention, before any consideration of baseline predictors. It was the number of responders in the treatment groups divided by the total number in the treatment groups, expressed as a percentage.

Predictive variables identified by the multivariate modelling were formulated into “tests” that represented different numbers of variables present

i.e. 1 of 6; 2 of 6; 3 of 6; 4 of 6; 5 of 6; 6 of 6. Numbers of participants positive for each test; numbers of participants negative for each test; numbers of participants who were responders to treatment; and numbers of participants, who were non-responders to treatment, were entered into two-by-two contingency tables. When cells in the 2X2 contingency tables had 0 values, LRs were estimated using a substitution formula (0.5 was added to each cell prior to calculations). Accuracy statistics were calculated giving sensitivity, specificity, Likelihood ratio following a positive test (LR+), and likelihood ratio following a negative test (LR-) for each test condition. These results were used to calculate post-test probability using the following method (Cleland, 2005).

Step 1 Pre-test odds = pre-test probability/1-pre-test probability

Step 2 Pre-test odds X LR = post-test odds

Step 3 Post-test odds/post-test odds +1 = post-test probability

The most useful tests are those which cause the greatest shift in post-test probability, in addition to consideration of the proportion of participants who meet the test criteria (Furukawa et al., 2008). Numbers and proportions of participants in each test group were recorded.

Post-test probability was calculated for the primary analysis of predictors of success at one year.

Combinations of predictor variables and post-test probability of success at one year, for analysis stratified by sex

Post-test probability was calculated using the same method of analysis, stratified by sex.

Analysis to investigate if specific combinations of named predictors alter post-test probability of success

The analysis described above gives post-test probability for a given *number* of predictors. This is conventionally the method for presenting Clinical Prediction Rules (CPRs). It does not calculate probability for groups of *specific* or named

predictors. It is likely that some variables may be stronger predictors than others and therefore post-test probability may change depending on the specific variables included in the models. A novel approach was employed in this study to address this issue. The multivariate models identified in the primary analysis were subjected to an additional post-estimation test that calculated adjusted predictions, in this case probabilities of success, for different named groups of predictors. It should be noted that this estimation of probability does not incorporate sensitivity and specificity of tests.

5.8.6. Secondary analyses

Analysis to investigate if model predicts outcome following physiotherapy intervention or natural course of disease

Prognostic models identified by the preceding analyses could predict response following physiotherapy attributable to the intervention, or might alternatively represent change as a result of natural course of the disease. As the MOA trial was an RCT, there was a control group (n = 28) who received usual care, with no physiotherapy intervention. The number in this group was small, lacking power to detect significant findings. However, analysis may give some indication of the effect of treatment versus no-treatment. Data analysis was performed as for the primary analysis.

Models derived from primary analyses were applied to the usual care group to see if the data fit.

Nine week treatment response as a predictor of long-term outcome

Success at nine weeks was investigated as a possible predictor of long-term success at one year. Usual clinical practice involves patient reassessment on completion of a treatment programme in order to evaluate treatment response and achievement of patient goals. Similarly the design of the clinical trial involved follow-up assessment at nine weeks to determine outcome post-

intervention. The same procedure was followed as for the one year outcome. Outcome was measured with WOMAC and GRC, with success determined by application of the OMERACT-OARSI responder criteria (Pham et al., 2004). Logistic regression was performed to determine a univariate association with successful outcome. Accuracy statistics and post-test probability were calculated. Stratified analysis was performed as for the primary analysis.

Additionally post-test probability was calculated for the cumulative effect of combining nine week treatment response with the optimal baseline predictor model. This would represent further gain of prognostic information and identify any improvement in prediction of one year success. The novel analysis was repeated for nine week response plus named baseline predictors to examine variations in probability with specific combinations of variables.

Baseline predictors of poor outcome at one year following physiotherapy intervention

Another research question is whether baseline variables can predict poor outcome at one year for participants with knee OA following physiotherapy intervention. Non-response at one year is not the same as having a poor outcome or deteriorating. Some participants may experience improvement at one year without reaching the threshold for success defined by the OMERACT-OARSI responder criteria. Therefore a separate analysis was required to differentiate participants who were not successful at one year from those who had experienced a deterioration, or “poor outcome” following physiotherapy intervention. MCID in WOMAC scores for patients with lower extremity OA was reported as 17% to 22% for improvement, and 6% to 22% for worsening (Angst et al., 2002).

Additionally, distribution of change scores was symmetrical for the pain and function subscales (Angst et al., 2002). These similarities suggest that inversion of OMERACT-OARSI responder criteria could be used to determine the threshold for poor outcome at one year when applied to the same outcome measures

(WOMAC and GRC). Deterioration of 20% applied to GRC would equate to change of greater than two steps resulting in a cut-point of “somewhat worse” (Appendix M).

Table 5.4 shows the responder criteria used to determine poor outcome at one year for participants with knee OA following physiotherapy intervention. Univariate and multivariate analyses were performed as for the primary analysis to identify baseline predictors.

Table 5.4 OMERACT-OARSI responder criteria to determine poor outcome at one year for participants with knee osteoarthritis following physiotherapy intervention

<p>≥ 50% deterioration in pain or function</p> <p>AND</p> <p>absolute deterioration ≥ 20</p>	<p>Deterioration in at least 2 of the following 3 scores:</p> <ul style="list-style-type: none"> • Pain ≥ 20% <p>AND absolute change ≥ 10</p> <ul style="list-style-type: none"> • Function ≥ 20% <p>AND absolute change ≥ 10</p> <ul style="list-style-type: none"> • Patient’s global assessment ≥ 20% <p>AND absolute change ≥ 10</p>
---	--

A diagrammatic representation of primary and secondary analyses is provided in Figure 5.0

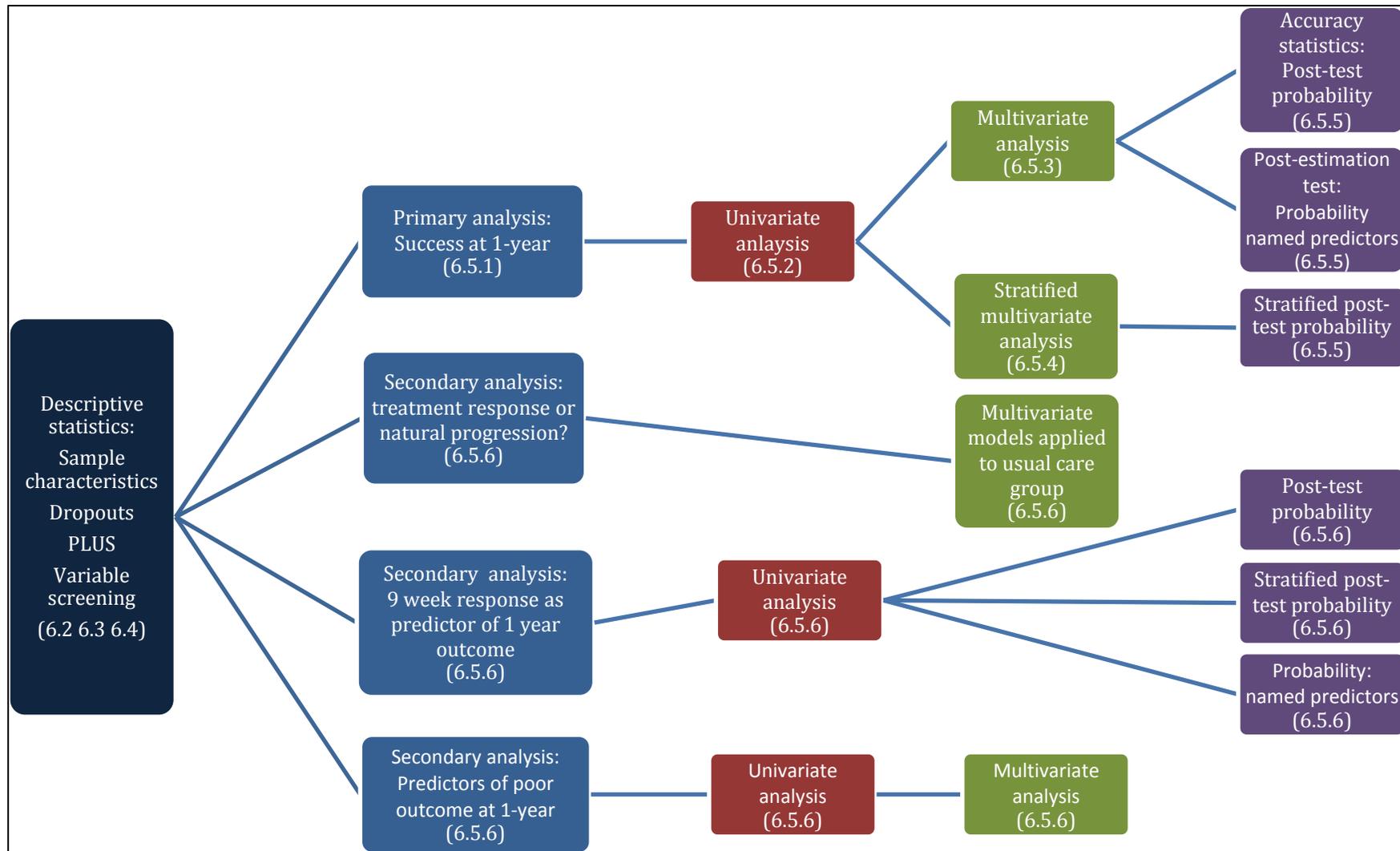


Figure 5.0 Overview of analysis (with reference to results section)

6. Study to identify predictors of outcome: Results

6.1. Overview

Reporting of results will follow the data analysis plan as outlined in the previous chapter.

Information on participant recruitment and retention will be presented, as well as descriptive results for the study sample. Baseline variables retained for analysis following the screening process are reported, with associated inter-rater reliability scores. Then one year outcome results are given, including *pre-test* probability of success.

Results pertinent to the primary research question that identify baseline predictors of success at one year following physiotherapy intervention will be presented, followed by results from stratified multivariate analysis. The optimal number of identified variables for predicting outcome and associated change in post-test probability of success will be reported. Specific named combinations of predictors and their associated probabilities of success are presented.

Results of secondary analyses follow. Application of identified models to usual care group data will examine whether they predict response to treatment, or outcome at one year due to natural course of knee osteoarthritis (OA). Nine week response as a predictor of outcome at one year will then be reported. Predictors of poor outcome at one year will be presented.

Finally, sensitivity analyses will be considered.

6.2. Characteristics of knee OA cohort from the clinical trial

Recruitment and retention of participants with knee OA in the Management of Osteoarthritis (MOA) trial is shown in Figure 6.1.

Characteristics of the trial cohort are shown in Table 6.1.

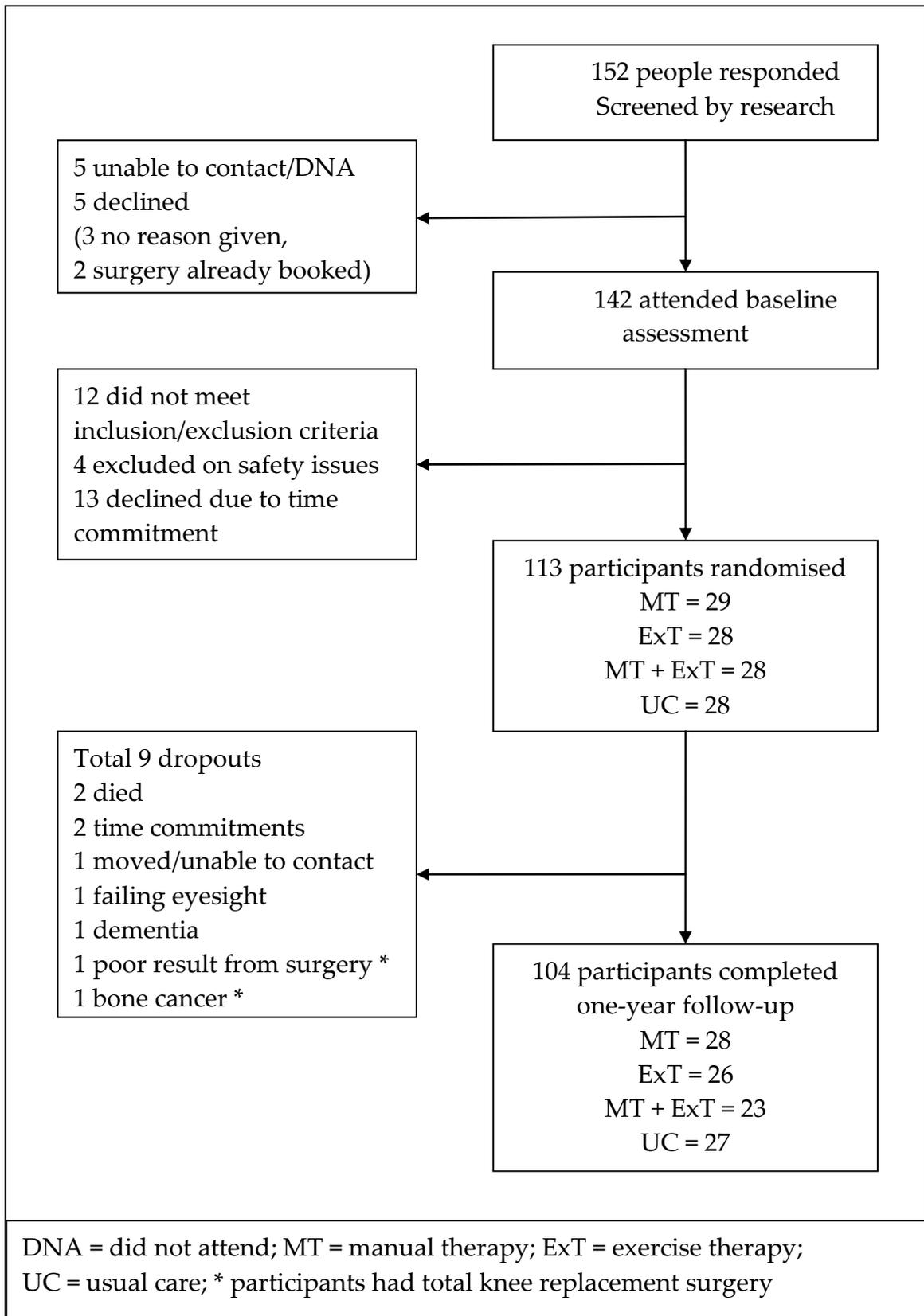


Figure 6.1 Flow chart showing study recruitment and retention of participants with knee osteoarthritis in the clinical trial.

Table 6.1 Characteristics of participants with knee osteoarthritis enrolled in the clinical trial

Participant Characteristic	All participants (n = 113)	Treatment (n = 85)	UsualCare (n = 28)
Age (in years)			
Mean (SD)	66.7 (9.6)	66.6(8.9)	67.0(11.7)
Range	37 - 92	45 – 87	37 -92
Sex			
Male	57 (50%)	40 (47%)	17 (61%)
Female	56 (50%)	45 (53%)	11 (39%)
Ethnicity*			
NZ/European	109 (96%)	84 (99%)	28 (100%)
Maori	10 (9%)	5 (6%)	1 (4%)
Samoan	1 (1%)	1 (1%)	0 (0%)
Marital Status			
Married/living with partner	77 (68%)	64 (75%)	13 (46%)
Divorced/widowed/never married	36 (32%)	21 (25%)	15 (54%)
Level of Education			
< 3 years high school	40 (36%)	32 (38%)	8 (29%)
>3 years high school	38 (34%)	27 (32%)	11 (39%)
Tertiary education/ training	34 (30%)	26 (31%)	9 (32%)
Employment status			
Working (full/part time)	43 (38%)	35 (41%)	10 (36%)
Not working (health reasons)	9 (8%)	16 (19%)	1 (4%)
Retired/unemployed/homemaker	61 (54%)	44 (52%)	17 (61%)
Duration of knee symptoms			
< 1 year	10 (9%)	8 (9%)	5 (18%)
1 – 2 years	20 (18%)	16 (19%)	4 (14%)
>2 – 5 years	31 (27%)	29 (34%)	8 (29%)
>5 – 10 years	21 (19%)	16 (19%)	8 (29%)
>10 years	31 (27%)	16 (19%)	3 (11%)
Mean (SD)	3.4 (1.3)	3.2 (1.2)	3.0 (1.3)
Baseline WOMAC (max = 240)			
Mean (SD)	96.3 (54.6)	98.8(56.1)	88.9(49.7)
Range	3 - 223	3 -223	3 - 194
Source of referral			
Orthopaedic/Rheumatology Clinic	45 (40%)	34 (40%)	11 (39%)
General Practice/Community	56 (50%)	40 (48%)	16 (57%)
Not recorded	12 (10%)	11 (13%)	1 (4%)
* participants can identify with more than one ethnic group			

Baseline Western Ontario and McMaster Universities Osteoarthritis index (WOMAC) scores are shown in Figure 6.2. Maximum possible score is 240, with higher scores indicating greater severity of symptoms. The range exhibited by the treatment cohort was 3 – 223, mean 96.3, standard deviation (SD) 54.6. The graph shows distribution was skewed slightly to the left, indicating mild disease was more frequent. However, a normal distribution was not required for selected statistical tests. Overall the graph shows there was a wide range in severity of symptoms at baseline, covering the full spectrum of knee OA disease.

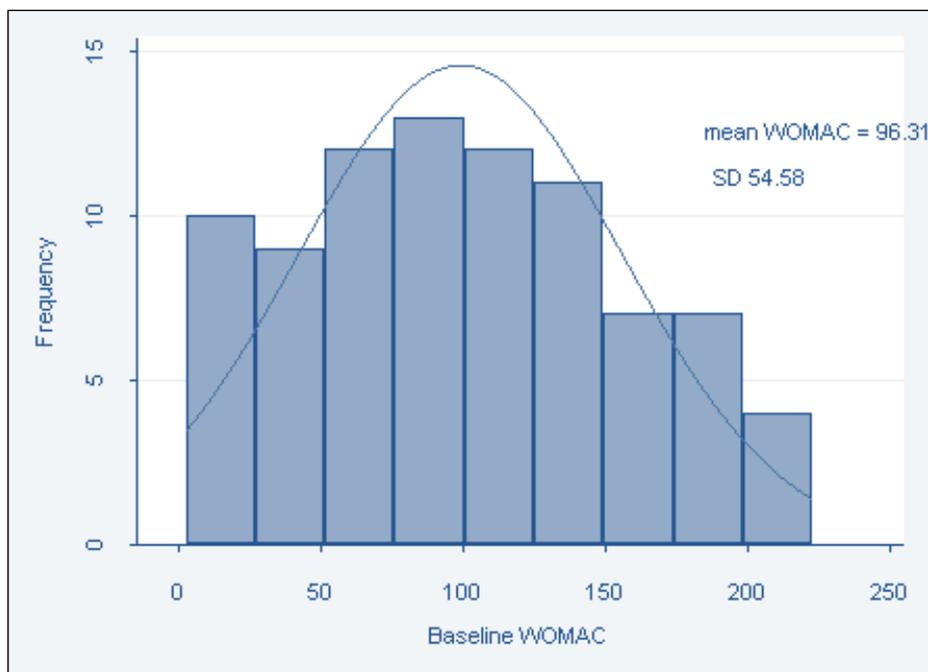


Figure 6.2 Distribution of baseline WOMAC scores for participants with knee osteoarthritis randomised to the treatment cohort in the clinical trial

Data for WOMAC subscales for pain and function at baseline and one year follow up are given in Appendix O, in addition to one year GRC scores. WOMAC scores were converted to a scale of 1 – 100 to allow for calculation of response using OARSI responder criteria. Details of WOMAC scores by group allocation are shown in Table 6.2

Table 6.2 Baseline and one year WOMAC pain and function subscale scores for treatment and usual care groups

	Baseline			One year follow up		
	WOMAC Pain subscale	WOMAC Function subscale	WOMAC Total	WOMAC Pain subscale	WOMAC Function subscale	WOMAC Total
	Mean (SD) Range	Mean (SD) Range	Mean (SD) Range	Mean (SD) Range	Mean (SD) Range	Mean (SD) Range
Treatment Group (n = 85)	41.2 (23.0) 0 – 92.0	40.4 (24.7) 0 – 93.5	41.3 (23.7) 1.3 – 92.9	27.3 (20.7) 0 – 76.0	28.0 (20.8) 0 – 80.0	28.1 (19.8) 0 – 72.9
Usual Care Group (n = 28)	35.2 (22.3) 0 – 86.0	36.3 (20.9) 1.2 – 75.3	37.1 (20.7) 1.3 - 80.8	31.3 (22.8) 0 – 86.0	33.4 (23.5) 1.8 – 89.4	33.4 (22.5) 2.9 – 83.3
WOMAC = Western Ontario and Macmaster Universities Osteoarthritis Index; SD = standard deviation						

6.3. Dropouts, missing data and surgical participants

Nine knee participants dropped out of the MOA trial before one year, giving a drop-out rate of 8% for these participants. Reasons for dropouts are shown in Figure 6.1. Two participants had total knee replacement (TKR) surgery but also withdrew from the trial: one due to poor outcome from surgery and one due to bone cancer. They were counted as non-success with physiotherapy intervention. One participant did not fill out the questionnaires at the final visit resulting in missing data for WOMAC and Global Rating of Change (GRC).

Demographic data for the drop-outs are shown in Table 6.3. There was a difference in WOMAC scores and level of education between participants remaining in the trial and the dropouts. Dropouts had significantly higher WOMAC scores ($p = 0.01$) indicating higher self-reported disability, and a significantly higher proportion had less than three years high school education ($p < 0.0005$), or more than three years high school education with no tertiary training ($p = 0.02$).

Nine participants did not attend for nine week follow-up, four of whom had withdrawn from the trial. Missing data from both nine week follow-up and one year follow-up were replaced using multiple imputations (MI).

Seventeen participants (15%) underwent major joint arthroplasty (including knee replacement, knee hemi-arthroplasty and hip replacement), and were considered as non-responders to physiotherapy intervention.

Table 6.3 Characteristics of completers and dropouts

Participant Characteristic	Completers (n = 104)	Dropouts (n = 9)	p-level
Age (in years)			
Mean (SD)	66.6 (9.6)	68.2 (9.7)	p = 0.63
Range	37 - 92	55 - 82	
Sex			
Male	50 (49%)	7 (78%)	p = 0.1
Female	54 (51%)	2 (22%)	
Ethnicity*			
NZ/European	100 (96%)	9 (100%)	p = 0.54
Maori	10 (9%)		
Samoan	1 (1%)		
Marital Status			
Married/living with partner	70 (67%)	7 (78%)	p = 0.5
Divorced/widowed/never married	34 (33%)	2 (22%)	
Level of Education			
< 3 years high school	32 (31%)	8 (89%)	p < 0.0005
> 3 years high school	38 (37%)	0 (0%)	p = 0.02
Tertiary education/training	33 (32%)	1 (11%)	p = 0.19
Employment status			
Working (full/part time)	39 (37%)	4 (44%)	p = 0.68
Not working (health reasons)	9 (9%)	0 (0%)	p = 0.35
Retired/unemployed/homemaker	56 (54%)	5 (56%)	p = 0.91
Duration of knee symptoms			
< 1 year	10 (10%)	1 (11%)	p = 0.88
1 – 2 years	18 (17%)	1 (11%)	p = 0.65
>2 – 5 years	28 (27%)	3 (33%)	p = 0.68
>5 – 10 years	21 (20%)	0 (0%)	p = 0.14
>10 years	27 (26%)	4 (44%)	p = 0.24
Mean (SD)	3.4 (1.3)	3.7 (1.3)	p = 0.51
Baseline WOMAC (max = 240)			
Mean (SD)	92.5 (53.4)	140.1 (51.9)	p = 0.01
Range	3 - 208	77 - 223	
Source of referral			
Orthopaedics/Rheumatology	39 (38%)	6 (67%)	p = 0.08
General Practice/Community	53 (51%)	3 (33%)	p = 0.3
Not reported	12 (12%)		
* participants can identify with more than one ethnic group			

6.4. Baseline variables for prognostic study

6.4.1. Screening of baseline variables

Variables retained for univariate analysis following screening procedures are shown in Table 6.4. Excluded variables and reasons for exclusions are summarised in Table 6.5. The term categorical has been used to describe some data, some (but not all) of which is ordinal. Inter-rater reliability values of retained clinical measures are also shown.

Table 6.4 Baseline variables included in univariate analyses: inter-rater reliability

Baseline Variable	Inter-rater reliability Kappa (% agreement) or ICC (95% CI, SEM)*	Inter-rater reliability†	Data type
Age <60; 60 to < 70; 70 to < 80; 80+ (< 57 years: ≥ 57 years)			Continuous Categorical‡ Binary §
Sex			Binary
BMI(kg/m ²) < 18.5; 18.5 to 24.99; 25 to 29.99; ≥ 30			Continuous Categorical‡
Self-efficacy/ catastrophizing/ fear-avoidance (PBSI)			Continuous
Depression (Two-item case finding instrument)			Binary
Smoking			Categorical
Physical activity			Categorical‡
Symptom duration (< 5yrs: ≥ 5 years)			Binary §
Previous knee injury	$\kappa = 0.62$ (91.3%)(Peat et al., 2003)	Substantial	Binary
Previous lower limb injury			Binary
History of falls			Binary
Previous helpful physiotherapy			Binary
Use of walking aid (Y/N)	$\kappa = 0.85$ (93.1%)(Peat et al., 2003)	Excellent	Binary §
Morning stiffness ≤ 30mins	$\kappa = 0.79$ (98.3%)(Peat et al., 2003) $\kappa = 0.58$ (95%CI, 0.38, 0.79) (Jones et al., 1992)	Substantial Moderate	Binary

Baseline Variable	Inter-rater reliability Kappa (% agreement) or ICC (95% CI, SEM)*	Inter-rater reliability†	Data type
NPRS	ICC 0.74 (SEM 0.66) - ICC 0.90 (SEM 0.68) (Mawdsley et al., 2002)	Substantial - Excellent	Continuous
Unilateral v bilateral pain (Y/N)	$\kappa = 0.72$ (86.2%)(Peat et al., 2003)	Substantial	Binary §
Anterior knee pain			Binary
Posterior knee pain			Binary
Medial knee pain			Binary
Lateral knee pain			Binary
Ipsilateral hip pain			Binary
Pain disturbs sleep (Y/N)	$\kappa = 0.79$ (86.2%)(Peat et al., 2003)	Substantial	Binary §
Knee locking	$\kappa = 0.71$ (91.4%)(Peat et al., 2003) $\kappa = 0.44$ (95%CI, 0.26 – 0.62) (80%) (Dervin et al., 2001)	Substantial Moderate	Binary
Self-report of instability from KOS-ADLS (Y/N)	ICC 0.72 (Fitzgerald et al., 2004a)	Moderate	Binary §
Knee swelling (Y/N)	$\kappa = 0.57$ (79.3%)(Peat et al., 2003) $\kappa = 0.33$ (95%CI, 0.17 – 0.49) (69%) (Dervin et al., 2001)	Moderate Fair	Binary §
Heat (Y/N)			Binary §
Irritability (the i-score)	ICC 0.64 (95% CI, 0.36 – 0.82) (SEM 2.57)	Moderate	Continuous
Valgus alignment (Y/N)	$\kappa = 0.30$ (95%CI, 0.21-0.64) (51.72%)	Fair	Binary §
Varus alignment (Y/N)	$\kappa = 0.30$ (95%CI, 0.21-0.64) (51.72%)	Fair	Binary §
Longitudinal arch angle (LAA) (Y/N)	$\kappa = 0.60$ (95% CI, 0.35-0.82)(86.2%)	Moderate	Binary §

Baseline Variable	Inter-rater reliability Kappa (% agreement) or ICC (95% CI, SEM)*	Inter-rater reliability†	Data type
Knee flexion ROM (degrees)	ICC 0.90 (Watkins et al., 1991) ICC 0.97 (SEM 3.9)(Fritz et al., 1998) ICC 0.98 (Brosseau et al., 2001)	Good Good Good	Continuous
Knee extension ROM (degrees)	ICC 0.86 (Watkins et al., 1991) ICC 0.94 (SEM 1.7)(Fritz et al., 1998) ICC 0.89 – 0.93 (Brosseau et al., 2001)	Good Good Good	Continuous
Hip flexion ROM (degrees)	ICC 0.80 (95% CI, 0.60-0.91) (SEM 8.6)	Good	Continuous
Ankle dorsi-flexion ROM (degrees)	ICC 0.50 (Elveru et al., 1988)	Moderate	Continuous
Thomas Test, iliopsoas restriction(degrees)	ICC 0.71 (95% CI, 0.47- 0.85) (SEM 6.69)	Moderate	Continuous
Thomas Test, rectus femoris restriction (degrees)	ICC 0.60 (95% CI, 0.31- 0.79) (SEM 12.2 degrees)	Moderate	Continuous
90/90 test, > 30 degrees from full knee extension – Y/N	ICC 0.58 (95% CI, 0.28-0.78)(SEM 7.58)	Moderate	Binary §
Tibiofemoral PA	$\kappa = 0.63$ (95% CI, 0.18-1.0) (93.1%)	Substantial	Categorical
Talocrural AP	$\kappa = 0.41$ (0.14, 0.55) (75.0%)	Moderate	Categorical
Quadriceps strength @ 90 degrees (kg)	ICC 0.72 (95%CI, 0.48- 0.86)(SEM 3.51)	Moderate	Continuous
Hamstring strength @ 0 -20 degrees (kg)	ICC 0.68 (95%CI, 0.25-0.86) (SEM 2.8)	Moderate	Continuous
Hip flexor strength (kg)	ICC 0.74 (95%CI, 0.48-0.87) (SEM 2.13)	Moderate	Continuous
Hip abductor strength (kg)	ICC 0.65 (95%CI, 0.38- 0.82)(SEM 1.47)	Moderate	Continuous
Hip external rotator strength (kg)	ICC 0.72 (95%CI, 0.48- 0.86)(SEM 1.23)	Moderate	Continuous
Hip internal rotator strength (kg)	ICC 0.71 (95%CI, 0.47- 0.85)(SEM 1.28)	Moderate	Continuous
McMurray's Test	$\kappa = 0.16$ (95%CI, - 0.01 – 0.33)(59%) (Dervin et al., 2001)	Slight	Binary

Baseline Variable	Inter-rater reliability Kappa (% agreement) or ICC (95% CI, SEM)*	Inter-rater reliability†	Data type
Knee valgus instability/stress test (Y/N)	Valgus instability $\kappa = 0.05$ (95%CI, - 0.03 – 0.23) (92%) (Dervin et al., 2001) Valgus stress test $\kappa = 0.36$ (lower 99%CL, 0.07) (67.9%) (Wood et al., 2006)	Slight Fair	Binary §
Knee varus instability/stress test (Y/N)	Varus instability $\kappa = 0$ (95%CI, - 0.18 – 0.18) (93%) (Dervin et al., 2001) Varus stress $\kappa = 0.24$ (lower 99%CL, -0.12) (76.8%) (Wood et al., 2006)	Poor Fair	Binary §
Patellar compression test	$\kappa = 0.44$ (0.03, 0.85) (82.76%)	Moderate	Binary
SPW time (seconds)	ICC 0.95 (95%CI, 0.90- 0.98)(SEM 2.00)	Good	Binary §
TUG time (seconds)	ICC 0.87 (95%CI, 0.74- 0.94)(SEM 0.84)	Good	Continuous
Sit – Stand (counted)	ICC 0.81 (95%CI, 0.63-0.91) (SEM 1.27)	Good	Binary §
Step-to-stool (counted)	ICC 0.91 (95%CI, 0.82-0.96) (SEM 5.8)	Good	Binary §

Baseline Variable	Inter-rater reliability Kappa (% agreement) or ICC (95% CI, SEM)*	Inter-rater reliability†	Data type
<p>ICC = inter class correlation coefficient; 95%CI = 95% confidence interval; SEM = standard error of measurement; * inter-rater reliability results from current investigation unless otherwise stated; †level of agreement: ICC >0.75 = good, 0.5 – 0.75 = moderate, < 0.5 = poor (Portney and Watkins, 2000); †level of agreement: κ: 0.81 – 1 = excellent, 0.61 – 0.8 = substantial, 0.41 – 0.6 = moderate, 0.21 – 0.4 = fair, 0.0 – 0.2 = slight, < 0.0 = poor (Landis and Koch, 1977); ‡ordered categorical; § modified with cut-point; BMI = body mass index; PBSI = pain beliefs and self efficacy instrument (Sandborgh et al., 2007); κ = kappa statistic; two questions combined; NPRS = numerical rating pain scale(Mawdsley et al., 2002); KOS-ADLS = Knee Outcome Survey-Activities of Daily Living Scale (Irrgang et al., 1998) ; ROM = range of movement; PA = postero-anterior accessory movement; AP = antero-posterior accessory movement; SPW = 40 metre self-paced walk (Kennedy et al., 2005); TUG = timed up and go test(Podsiadlo and Richardson, 1991)</p>			

Table 6.5 Reasons for excluding variables from analysis

Reason for exclusion	Variable
Variables required for hip OA study	Hip pain location Hip morning stiffness Duration hip symptoms Hip distraction test Tender on palpation greater trochanter Hip internal/external ROM
Variables used as inclusion criteria	Worse joint hip/knee Heat on palpation Bony enlargement Crepitus Joint line tenderness
Variables used to determine patient safety	All cardiovascular screening questions Lumbar spine ROM/pain
Non-discriminatory/ high prevalence	Crepitus Straight leg raise test
Two or more variables collapsed into one retained variable	Hip pain + worse hip = ipsilateral hip pain Previous hip injury + previous other leg injury = previous lower limb injury Irritability questions into the i-score Nights in last week disturbed sleep + number of times per night sleep disturbed = disturbed sleep Ever had physiotherapy treatment + was physiotherapy helpful = previous helpful physiotherapy treatment
Supplementary questions not required for analysis	All contralateral limb questions Pain associated with locking or giving way/instability Knee flexion/extension end feel and pain-resistance sequence Exertion and pain response data associated with PPMs
Duplicates (more than one variable examining same characteristic)	Alignment in sagittal plane (fixed flexion deformity/ hyperextension) examined as knee extension ROM Heat/swelling on palpation examined as history of heat/swelling
OA = osteoarthritis; ROM = range of motion; PPM = physical performance measure	

6.4.2. Creation of cut-points in data

A decision was made to avoid cutting data unnecessarily in order to retain as much information as possible. However in some cases creation of cut-points was justified as categories or dichotomised data made more clinical sense.

Continuous data

The following continuous variables were considered for creation of cutpoints:

- **Age.** Two methods of creating cut-points were employed for age. Receiver operating characteristics (ROC) curve analysis produced a cut-point of 57 years based on LR data (LR+ 1.13, LR- 0.23). An alternative frequently adopted for age is to use age-bands. Age was categorised into < 60years; 60 - <70 years; 70 - < 80 years; \geq 80 years.
- **BMI.** BMI was also investigated for a cut-point to dichotomise data using ROC curve analysis. However, the shape of the curve did not reveal any suitable cut-points, nor were any detected by examining LRs. Secondly, BMI was divided according to the World Health Organisation (WHO) classification: $<18.5 \text{ kg/m}^2$ = underweight; $18.5\text{--}24.99 \text{ kg/m}^2$ = normal; $\geq 25.0\text{--}29.99 \text{ kg/m}^2$ = overweight: and $\geq 30.0 \text{ kg/m}^2$ = obese (World Health Organisation, 2006).
- **Twenty-centimetre step-ups.** During assessments it was clear that many participants could complete the task with ease. Those who found the task difficult stopped early in the test or well short of fifty repetitions. This was reasonable justification to create a cut point, therefore it became:
“Can the participant complete 50 step ups on a 20cm step?” with the response Yes/No.

Cut versions of age and BMI are considered further in a sensitivity analysis (section 6.5.7).

Categorical data

- Knee alignment and foot posture remained as categorical variables for analysis.
- Instability, waking at night with knee pain, heat, swelling, valgus stress test, varus stress test, and use of walking aid were dichotomised into yes/no answers to better reflect usual clinical practice.
- Duration of knee symptoms. Participant response was simplified for analysis, creating a cut point of five years or more. Five years is an easy reference point for participants, in addition to being a time period that would allow for reasonable progression of disease.

6.5. Predictors of success at one year for participants with knee OA following physiotherapy intervention

6.5.1. Successful outcome at one year

At one year, using OMERACT-OARSI responder criteria, thirty-five participants had good outcome; seventy-eight participants did not meet the threshold for good outcome. The probability of a good outcome was 31% for the whole clinical trial cohort. Response data is given in Appendix P.

Eighty-five participants were randomised to treatment groups forming the treatment cohort. This cohort was used for analyses in the main prognostic study, except where otherwise stated. At one year, using OMERACT-OARSI responder criteria, thirty participants had good outcome, and were considered treatment responders. The remaining fifty-five participants did not reach the threshold for good outcome, and were considered non-responders to physiotherapy intervention. Non-responders included participants who underwent joint arthroplasty before the one year follow-up. Probability of success in the treatment cohort was 35% (Table 6.6).

Table 6.6 Responders at one year and probabilities of successful outcome.

Success at one year	All participants n = 113	Combined treatment groups n = 85	ExT n = 28	MT n = 29	ExT + MT n = 28	UC group n = 28
Number of responders	35/113	30/85	10/28	12/29	8/28	5/28
Probability of success	31%	35%	36%	41%	29%	18%
ExT = exercise therapy; MT = manual therapy; UC = usual care						

6.5.2. Univariate analysis identifying baseline variables associated with success at one year

Fifteen variables had a univariate association with outcome at one year ($p < 0.2$); these were retained for multivariate analysis. Significance levels and odds ratios (OR) are shown in Table 6.7. The table is ordered by p-value as this was the criterion used to retain variables for multivariate analysis. Of note are the modest ORs for knee pain, irritability, and pain beliefs/self efficacy which are close to 1.0 suggesting very weak association with outcome. In other words, odds of a successful outcome at one year are about the same regardless of pain, irritability, or pain beliefs scores. Variables with ORs < 1.0 are usually interpreted as having a protective effect. Analysis revealed previous knee injury with OR 0.43. However, in the context of this study it does not make sense to report previous knee injury “protects” against success. An alternative interpretation is *absence* of knee injury was associated with success. This interpretation will be used for the remainder of the primary analysis.

Table 6.7 Baseline variables associated with success at one year for participants with knee osteoarthritis following physiotherapy intervention: univariate analysis

Baseline variable	OR (95% CI)	p-value
Disturbed sleep	4.15 (1.47, 11.73)	0.004
Presence of posterior knee pain	3.05 (1.20, 7.73)	0.02
Stronger hamstring strength	7.22 (1.09, 47.78)	0.03
Positive patella compression test	2.55 (1.02, 6.37)	0.04
Higher reported knee pain	1.06 (0.99, 1.13)	0.07
Previous knee injury	0.43 (0.16, 1.17)	0.09
Knee locking	2.17 (0.87, 5.37)	0.09
Higher irritability	1.13 (0.98, 1.31)	0.09
20 cm step test (< 50)	2.11 (0.86, 5.23)	0.10
Knee instability	2.15 (0.83, 5.62)	0.12
Stronger hip external rotator muscle strength	16.74 (0.44, 629.62)	0.12
Poor pain beliefs/self efficacy	1.03 (0.99, 1.06)	0.13
Stronger hip flexor muscle strength	4.15 (0.59, 29.38)	0.15
Symptom duration (> 5 years)	1.94 (0.78, 4.79)	0.15
Female Sex	1.93 (0.77, 4.80)	0.15
OR = odds ratio; 95% CI = 95% confidence interval		

6.5.3. Multivariate analysis identifying baseline variables associated with success at one year

Logistic regression with backward elimination resulted in six variables being retained in the final model: presence of posterior knee pain, disturbed sleep, duration of symptoms (> 5years), knee instability, female sex, and *absence* of previous knee injury. Results ordered by p-value are shown in Table 6.8.

Symptom duration (> 5 years) was marginal for elimination from regression analysis ($p < 0.096$), with the *a priori* p-level set at 0.1. However, when it was removed the p-value for the model increased slightly to $p < 0.0004$ and Akaike (AIC) increased to 98.15, indicating a weaker model. Also, symptom duration is often incorporated into clinical assessment, giving further justification to retain it in the model.

Table 6.8 Predictors of success at one year for participants with knee osteoarthritis following physiotherapy intervention: multivariate analysis

Baseline Variable	Participants with predictor (%)	Odds ratio (95% CI)	p - value
Absence of previous knee injury	29/83 (35%)*	0.22 (0.06, 0.79)	$p < 0.02$
Presence of posterior knee pain	31/85 (36%)	3.65 (1.14, 11.69)	$p < 0.03$
Disturbed sleep	51/85 (60%)	3.42 (1.04, 11.28)	$p < 0.04$
Female sex	45/85 (53%)	3.23 (1.01, 10.35)	$p < 0.05$
Knee instability	25/85 (29%)	3.27 (1.01, 10.67)	$p < 0.05$
Symptom duration (> 5 years)	42/85 (49%)	2.64 (0.84, 8.25)	$p < 0.096$
$p < 0.0003$; AIC = 97.29; BIC = 114.22; GOF: $p = 0.62$; AUC = 0.80			
95% CI = 95% confidence interval; * Missing data 2 participants; AIC = Akaike information criterion; BIC = Bayesian information criterion; GOF = Hosmer Lemeshow goodness of fit; AUC = Area under the curve			

Post-estimation analysis revealed no collinearity between retained variables. Hosmer Lemeshow goodness of fit (GOF) p-value was 0.62, indicating no significant difference between the observed values and those predicted by the model. Therefore the model was a good fit for the data. Area under the curve (AUC) of 0.8 shows the model discriminates between success and non-success at one year.

6.5.4. Stratified analysis

Sex was retained for multivariate analysis and is one of the six variables in the final model. It was therefore decided to perform a stratified analysis by sex, recognising its potential as a strong confounder.

Multivariate analysis stratified by sex identifying baseline variables associated with success at one year

Logistic regression with backward elimination was repeated for males and females separately. For males, only one baseline variable was retained in the model as a significant predictor of success at one year, posterior knee pain. For females, *absence* of previous knee injury, instability, and disturbed sleep were retained as predictors of success at one year (Table 6.9).

Table 6.9 Predictors of success at one year stratified by sex

Baseline Variable	Odds ratio (95% CI)	p - value	AUC
Male*			
Posterior knee pain	8.44 (1.68, 42.39)	p < 0.01	0.63
Female†			
Previous knee injury‡	0.14 (0.03, 0.81)	p < 0.03	0.73
Instability	6.21 (0.95, 40.66)	p < 0.06	
Disturbed sleep	17.60 (2.34, 130.32)	p < 0.01	
*p < 0.0056; AIC = 40.65; BIC = 43.81; GOF – n/a;			
†p < 0.001; AIC = 52.97; BIC = 60.19; GOF p = 0.95			
95% CI = 95% confidence interval; AUC = area under the curve; ‡variable retained as <i>absence</i> of knee injury; AIC = Akaike information criterion; BIC = Bayesian information criterion; GOF = Hosmer Lemeshow goodness of fit			

Post-estimation analysis of the model for females revealed no collinearity between retained variables.

For females, Hosmer Lemeshow GOF p-value was 0.95, indicating no significant difference between the observed values and those predicted by the model, suggesting that the model was a good fit for the data.

For males, the AUC = 0.63 and for females AUC = 0.73, suggesting the models discriminate between success and non-success at one year.

6.5.5. Clusters of predictor variables, and post-test probability of success at one year

All six variables retained in the main multivariate model were dichotomous, allowing further analysis to take place. Clustering different numbers of variables together produced “tests”: i.e. one out of six, two out of six etc. The different

“test” conditions are shown in Table 6.10, with their associated effect on post-test probability.

Table 6.10 Different test conditions for baseline predictors with associated likelihood ratios and post-test probabilities

Variables present	Number of participants in each group (n = 83) * (%)	Sensitivity (95% CI)	Specificity (95% CI)	LR+ (95% CI)	LR- (95% CI)	Post-test probability of success with +’ve test	Post-test probability of success with -’ve test
≥ 1 of 6	10 (12.0%)	100% (88.4, 100)	1.9% (0.1, 10.1)	1.01 (0.95, 1.08)	0.58 (0.02, 13.8)	35.2%	23.8%
≥ 2 of 6	18 (21.7%)	100% (88.4, 100)	20.8% (10.8, 34.1)	1.25 (1.08, 1.45)	0.08 (0.0, 1.24)	40.2%	3.1%
≥ 3 of 6	30 (36.1%)	90% (73.5, 97.9)	47.2% (33.3, 61.4)	1.68 (1.27, 2.23)	0.24 (0.09, 0.67)	47.5%	11.4%
≥ 4 of 6	16 (19.3%)	56.7% (37.4, 74.5)	84.9% (72.4, 93.3)	3.59 (1.8, 7.15)	0.52 (0.34, 0.78)	65.9%	21.9%
≥ 5 of 6	9 (10.8%)	26.7% (12.3, 45.9)	98.1% (89.9, 100)	9.87 (1.84, 52.9)	0.75 (0.60, 0.93)	84.2%	28.8%
≥ 6 of 6	0 (0%)	0.0% (0.0, 11.6)	100% (93.3, 100)	1.74 (0.04, 85.6)	0.99 (0.94, 1.05)	48.4%	34.8%
Based on pre-test probability of success = 35%							
*Missing data for history of knee injury for 2 participants; 95% CI = 95% confidence interval; LR+ = likelihood ratio of positive test; LR- = likelihood ratio of negative test; +’ve test = positive test; -’ve test = negative test							

The greatest increase in post-test probability was the presence of ≥ 5 out of 6 variables with a LR+ of 9.87 (95% CI: 1.84, 52.9) giving a post-test probability of 84.2%. However, only nine participants (10.8%) met this test condition, therefore confidence intervals were wide with increased uncertainty in the result. When ≥ 4 out of 6 variables were present LR+ was 3.59 (95% CI: 1.8, 7.15) giving a post-test probability of 65.9%. A cumulative total of twenty-five participants (30.1%) met this test condition (19.3% + 10.8% + 0%). Consequently the 95% CI was narrower giving greater confidence in the result.

The strongest result producing a decrease in post-test probability was having less than 3 out of 6 variables, giving a post-test probability of 11.4%. This result stems from the LR- of 1.68 (95% CI: 1.27, 2.23) for the ≥ 3 out of 6 test condition. The proportion of participants exhibiting less than three predictive variables was 33.7% (21.7% + 12.0%).

When one or less variables were present the LR- (for ≥ 2 out of 6) was 0.08 (95% CI: 0.0, 1.24) and the post-test probability dropped to 3.1%. However, as the 95% CI crosses 1.0, this result is not-significant. Additionally only 12% of participants met this test criteria.

Combinations of predictor variables and post-test probability of success at one year, for analysis stratified by sex

All variables retained in the analyses stratified by sex were dichotomous allowing for calculation of accuracy statistics and post-test probability of success. Results are shown in Table 6.11.

Table 6.11 Different test conditions for baseline predictors with associated likelihood ratios and post-test probabilities stratified by sex

Variables present	Number of participants in each group (n = 85) (%)*	Sensitivity (95% CI)	Specificity (95% CI)	LR+ (95% CI)	LR- (95% CI)	Post-test probability of success with +’ve test	Post-test probability of success with -’ve test
Males (n = 40)							
Posterior knee pain	15 (37.5%)	72.7% (39, 94)	75.9% (56.5, 89.7)	3.01 (1.44, 6.31)	0.36 (0.13, 0.96)	61.8%	16.2%
Females (n = 45)							
≥ 1 of 3	15 (33.3%)	100% (81.5, 100)	18.5% (6.3, 38.1)	1.21 (0.99, 1.48)	0.13 (0.01, 2.28)	39.5%	6.5%
≥ 2 of 3	23 (51.1%)	88.9% (65.3, 98.6)	66.7% (46.0, 83.5)	2.56 (1.48, 4.42)	0.20 (0.06, 0.65)	57.9%	9.7%
≥ 3 of 3	2 (4.4%)	11.1% (1.4, 34.7)	100% (87.2, 100)	7.37 (0.37, 145)	0.88 (0.74, 1.06)	79.9%	32.1%
Based on pre-test probability of success = 35%; *5 participants had 0 of 3 variables (11.1%)							
95% CI = 95% confidence interval; LR+ = likelihood ratio of positive test; LR- = likelihood ratio of negative test; +’ve test = positive test; -’ve test = negative test							

Males with posterior knee pain had increased probability of success at 62%. The absence of posterior knee pain decreased probability of success to 16%, less than half the pre-test probability of success (35%). Posterior knee pain was present in 37.5% of male participants in the treatment cohort.

The greatest increase in post-test probability for females was ≥ 2 out of 3 variables (absence of knee injury, instability, disturbed sleep), with a LR+ 2.56 (95% CI: 1.48, 4.42) giving 57.9% probability of success. There were twenty-five female participants with two or more variables (55.5%), with a narrow 95% CI increasing confidence in the result.

When one or less variables were present the LR- (for ≥ 2 out of 6) was 0.2 (95% CI: 0.06, 0.65) and the post-test probability dropped to 9.7%. The proportion of female participants exhibiting one or less variables was 44.4% (33.3% + 11.1%).

Probability of success at one year given specific combinations of named predictors

The novel analysis to identify probability of success given specific sets of named predictors gave a range of probabilities (Table 6.12). The preceding analysis identified the model with the greatest post-test probability of success is four or more predictors. Therefore, probabilities of success are reported for variants of that model with four or five predictors. There is no report for a model with all six predictors as there were no participants in that category. Similarly some combinations of four or five variables did not have any participants; therefore probability is not reported (e.g. female, absence of knee injury, duration of symptoms (> 5 years) with posterior knee pain). Results in Table 6.12 are in order of increasing probability.

Table 6.12 shows the range of probability for success is 55% to 68% when 4 out of 6 predictors were present. A female participant with disturbed sleep, posterior knee pain, and a longer duration of symptoms (> 5 years) had increased probability of success from 35% (pre-test probability) to 55%. However, a female

participant with a different set of three predictors, disturbed sleep, posterior knee pain, and no previous knee injury, had a greater probability of success of 68%. Similarly, probability of success ranged from 80% to 87% when 5 out of 6 predictors were present.

Table 6.12 Model variants showing specific combinations of named predictors and range of probability for success at one year

Model	Disturbed sleep	Posterior knee pain	Absence of previous knee injury	Symptom duration > 5 years	Instability	Female sex	Probability (95% CI)
4 out of 6	✓	✓		✓		✓	0.55 (0.26, 0.81)
	✓	✓		✓	✓		0.56 (0.23, 0.84)
			✓	✓	✓	✓	0.60 (0.23, 0.88)
	✓		✓	✓		✓	0.61 (0.32, 0.84)
	✓		✓	✓	✓		0.61 (0.30, 0.85)
	✓	✓	✓	✓			0.64 (0.32, 0.87)
	✓		✓		✓	✓	0.66 (0.34, 0.88)
	✓	✓	✓		✓		0.68 (0.33, 0.91)
	✓	✓	✓			✓	0.68 (0.37, 0.89)
5 out of 6	✓	✓		✓	✓	✓	0.80 (0.43, 0.96)
	✓	✓	✓	✓	✓		0.85 (0.56, 0.96)
	✓	✓	✓	✓		✓	0.85 (0.58, 0.96)
	✓	✓	✓		✓	✓	0.87 (0.58, 0.97)
Note: no participants with 6 out of 6 predictors							

6.5.6. Secondary analyses

Predictors of outcome following intervention or predictors of natural course of knee OA

Pre-test probability of success at one year in the usual care group was 18% (Table 6.6). Multivariate regression could not be performed in the usual care group as there were fifteen variables and only twenty-eight participants.

An alternative analysis involved forcing a treatment into the analysis and retaining it regardless of p-value. The treatment variable was re-coded and re-named prior to further analysis, making usual care the predictor variable. The result of forcing the usual care variable (yes/no) into the analysis was a four variable model: presence of posterior knee pain, previous knee injury, disturbed sleep; and usual care (Table 6.13). Previous knee injury and usual care returned small ORs. As previously, *absence* of knee injury was retained as a predictor of success. The small OR means usual care was associated with *non-success*. The model is different to that which predicts success in the treatment cohort, although three variables are common to both models.

Table 6.13 Predictors of success at one year when treatment (yes/no) forced into multivariate analysis

Baseline Variable	Odds ratio (95% CI)	p - value
Posterior knee pain	3.75 (1.48, 9.48)	p < 0.05
Previous knee injury*	0.43 (0.16, 1.13)	p < 0.09
Disturbed sleep	2.38 (0.92, 6.19)	p < 0.07
Usual care (No treatment)	0.39 (0.12, 1.23)	p < 0.11
p < 0.0007; AIC = 129.04; BIC = 142.59; GOF p = 0.04; AUC = 0.75		
95%CI = 95% confidence interval; *variable retained as <i>absence</i> of knee injury; AIC = Akaike information criterion; BIC = Bayesian information criterion; GOF = Hosmer Lemeshow goodness of fit; AUC = area under the curve		

Post-estimation analysis revealed no collinearity between retained variables. Hosmer Lemeshow GOF p-value was 0.04, indicating significant difference between the observed values and those predicted by the model. Therefore the model did not fit the data. AUC of 0.8 shows the model discriminates between success and non-success at one year.

Another method used to investigate whether the six-variable model predicted success at one year as a response to treatment or predicts natural course, was to apply the model to the usual care group data. Results are shown in Table 6.14. Four participants (11.3%) in the usual care group exhibited at least 4 out of 6 baseline predictors, compared to 30.1% in the treatment cohort. None of the LRs was significant, with all the 95% confidence intervals crossing 1.0 (Table 6.14).

Table 6.14 Different test conditions for baseline predictors with associated likelihood ratios and post-test probabilities using data from usual care group

Variables present	Number of participants in each group (%)*	Sensitivity (95% CI)	Specificity (95% CI)	LR+ (95% CI)	LR- (95% CI)	Post-test probability of success with +’ve test	Post-test probability of success with -’ve test
1 of 6	7 (25%)	0% (0, 52.2)	75% (55.1, 89.3)	0.32 (0.02, 4.91)	1.24 (0.9, 1.71)	6.6%	21.4%
2 of 6	10 (35.7%)	60% (14.7, 94.7)	69.6% (47.1, 86.8)	1.87 (0.76, 4.59)	0.61 (0.23, 1.62)	29.1%	11.8%
3 of 6	6 (21.4%)	40% (5.3, 85.3)	82.6% (61.2, 95)	2.22 (0.63, 7.84)	0.72 (0.36, 1.45)	32.8%	13.7%
4 of 6	1 (3.6%)	0% (0, 52.2)	95.7% (78.1, 99.9)	1.33 (0.06, 28.8)	1.0 (0.75, 1.27)	22.6%	17.7%
5 of 6	3 (10.7%)	0% (0, 52.2)	87% (66.4, 97.2)	0.57 (0.03, 9.63)	1.07 (0.8, 1.44)	11.1%	19.0%
6 of 6	0 (0%)	0% (0, 52.2)	100% (85.2, 100)	4 (0.09, 182)	0.94 (0.73, 1.2)	46.8%	17.1%
Based on pre-test probability of success = 18%; *1 patient had none of 6 variables (3.6%)							
95% CI – 95% confidence interval; LR+ = likelihood ratio of positive test; LR- = likelihood ratio of negative test; +’ve = positive test; -’ve = negative test							

Nine week response as a predictor of success at one year

At nine weeks, seventy-seven participants of a possible eighty-five in the treatment cohort returned for assessment following physiotherapy intervention. Multiple imputations (MI) data was available for WOMAC and GRC, allowing calculation of nine week outcome for all eighty-five participants. Approximately half the treatment cohort exceeded the threshold for successful response at nine weeks giving probability of success of 49% (Table 6.15). This compares with pre-test probability of success at one year of 35%.

Table 6.15 Success at nine weeks for participants with knee osteoarthritis following physiotherapy intervention

	Success at nine weeks (%)
Treatment cohort (n= 85)	42/85 (49%)
Males (n = 40)	22/40 (55%)
Females (n = 45)	20/45 (44%)
Usual care (n = 28)	2/28 (7%)

Results of univariate logistic regression show success at nine weeks significantly increases the odds of success at one year (OR 4.81). When stratified by sex, the odds increase for males (OR 14.17), but decrease for females (OR 3.86). AUC = 0.68 shows model is likely to discriminate between success and non-success. Further detail is shown in Table 6.16. Low numbers precluded analysis of nine week response as a predictor of outcome in the usual care group.

Table 6.16 Odds of success at one year following success at nine weeks: univariate analysis

	Odds ratio (95% CI)	p - value
nine week success	4.81 (1.81, 12.8)	p < 0.002
nine week success (male)	14.17 (1.59, 125.87)	p < 0.017
nine week success (female)	3.86 (1.11, 13.46)	p < 0.034
AUC = 0.68		
95%CI = 95% confidence interval; AUC = area under the curve		

Accuracy statistics and post-test probability for nine week response being a predictor of success at one year are shown in Table 6.17. Success at nine weeks increases the likelihood of success at one year from 35% to 52%. Failure to succeed at nine weeks decreases the likelihood of success at one year to 18%. These effects are greater in males, although the finding that failure to succeed at nine weeks decreases post-test probability of success to 8% (in males) should be regarded cautiously as the 95% CI crosses 1.0.

Investigation of the cumulative effect of the primary model of baseline predictors and nine week response, to success at one year produced a seven – variable model (Table 6.18).

When at least four out of seven predictors were present post-test probability increased from 35% to 55.8%, with nearly half the participants (48.2%) meeting this test condition. The increase in post-test probability was even greater for the test at least five out of seven predictors (86.1%), although less participants met this condition (21.7%) resulting in wider 95% CI.

The strongest result for decreased post-test probability was less than three out of seven predictors, reducing the likelihood of success to 6.5%, with 25.3% of participants meeting this test condition.

Table 6.17 Nine week response as a predictor of success at one year.

	Number of participants in each group (%)	Sensitivity (95% CI)	Specificity (95% CI)	LR+ (95% CI)	LR- (95% CI)	Post-test probability of success with +’ve test	Post-test probability of success with -’ve test
Nine week response	85 (100%)	73.3% (54.1, 87.7)	63.6% (49.6, 76.2)	2.02 (1.34, 3.04)	0.42 (0.22, 0.78)	52%	18%
Nine week response (males)	40 (47%)	90.9% (58.7, 99.8)	58.6% (38.9, 76.5)	2.2 (1.37, 3.52)	0.16 (0.02, 1.03)	54%	8%
Nine week response (females)	45 (53%)	63.2% (38.4, 83.7)	69.2% (48.2, 85.7)	2.05 (1.05, 4.02)	0.53 (0.28, 1.01)	52%	22%
Based on pre-test probability of success = 35%							
95% CI – 95% confidence interval; LR+ = likelihood ratio of positive test; LR- = likelihood ratio of negative test; +’ve = positive test; -’ve = negative test							

Table 6.18 Probability of success at one year with baseline predictors AND success at nine weeks.

Variables present	Number of participants in each group n = 83*	Sensitivity (95% CI)	Specificity (95% CI)	LR+ (95% CI)	LR- (95% CI)	Post-test probability of success with +ve test	Post-test probability of success with -ve test
1 of 7	8 (9.6%)	100% (88.4, 100)	0% (0, 6.7)	0.99 (0.94, 1.05)	1.74 (0.04, 85.6)	34.8%	48.4%
2 of 7	13 (15.7%)	100% (88.4, 100)	15.1% (6.8, 27.6)	1.17 (1.03, 1.32)	0.10 (0.01, 1.72)	38.7%	5.1%
3 of 7	22 (26.5%)	96.7% (82.8, 99.9)	37.6% (24.8, 52.1)	1.53 (1.23, 1.92)	0.13 (0.03, 0.63)	45.2%	6.5%
4 of 7	22 (26.5%)	76.7% (57.7, 90.1)	67.9% (53.7, 80.1)	2.34 (1.52, 3.61)	0.34 (0.19, 0.69)	55.8%	16.2%
5 of 7	11 (13.3%)	53.3% (34.3, 71.7)	96.2% (87.0, 99.5)	11.5 (3.28, 40.3)	0.49 (0.34, 0.72)	86.1%	20.9%
6 of 7	7 (8.4%)	20% (7.7, 38.6)	98.1% (89.9, 100)	7.55 (1.35, 42.1)	0.81 (0.67, 0.98)	80.3%	30.1%
7 of 7	0 (0%)	0% (0.0, 11.6)	100% (93.3, 100)	1.74 (0.04, 85.6)	0.99 (0.94, 1.05)	48.4%	34.8%
Based on pre-test probability of success = 35%							
*2 patients with missing knee injury data; 95% CI – 95% confidence interval; LR+ = likelihood ratio of positive test; LR- = likelihood ratio of negative test; +ve = positive test; -ve = negative test							

Probability of success at one year given specific combinations of nine week response and named predictors

Repetition of the novel analysis employed earlier aimed to reveal the particular impact of including nine week response with baseline predictors on probability of success at one year. Results shown in Table 6.19 include only combinations with nine week response included.

Table 6.19 Nine week response and named predictors – probability of success at one year

Model	Nine week response	Disturbed sleep	Posterior knee pain	NO previous knee injury	Symptom duration > 5 years	Instability	Female sex	Probability (95% CI)
4 out of 7	✓			✓		✓	✓	0.16 (0.01, 0.76)
	✓		✓		✓		✓	0.33 (0.04, 0.83)
	✓	✓		✓		✓		0.43 (0.10, 0.84)
	✓	✓		✓			✓	0.45 (0.16, 0.78)
	✓	✓	✓				✓	0.51 (0.13, 0.88)
	✓	✓		✓	✓			0.60 (0.21, 0.90)
	✓	✓	✓		✓			0.66 (0.24, 0.92)
	✓	✓	✓	✓				0.68 (0.21, 0.94)
5 out of 7	✓	✓		✓	✓	✓		0.73 (0.28, 0.95)
	✓	✓		✓	✓		✓	0.74 (0.35, 0.94)
	✓	✓	✓		✓	✓		0.77 (0.28, 0.97)
	✓	✓	✓	✓		✓		0.79 (0.32, 0.97)
	✓	✓	✓	✓			✓	0.80 (0.42, 0.96)
	✓	✓	✓	✓	✓			0.88 (0.47, 0.98)
6 out of 7	✓	✓	✓		✓	✓	✓	0.86 (0.31, 0.99)
	✓	✓	✓	✓		✓	✓	0.89 (0.42, 0.99)
	✓	✓	✓	✓	✓	✓		0.93 (0.57, 0.99)
	✓	✓	✓	✓	✓		✓	0.93 (0.63, 0.99)
No participants with 7 out of 7 predictors								

Predictors of poor outcome at one year

The number of participants who met the criteria for poor outcome and the associated probability is shown in Table 6.20.

Table 6.20 Number of responders and probability of poor outcome at one year

	All participants n = 113	Combined treatment groups n = 85	Usual care group n = 28
Poor outcome at one year			
Number of responders	10/113	9/85	1/28
Probability of worsening	9%	11%	4%

Univariate logistic regression was performed to identify baseline variables associated with poor outcome at one year. Fourteen variables with $p < 0.2$ are shown in Table 6.21. Five of these (positive patellar compression test, greater hamstring strength, higher pain, catastrophizing/poor self efficacy and inability to do fifty step-ups) also had a univariate association with success at one year. It should be noted that for both outcomes (good/poor) ORs for pain, pain beliefs/self efficacy and 20cm step test, were close to 1.0, indicating weak association with either outcome.

Table 6.21 Baseline variables associated with poor outcome at one year: univariate analysis

Baseline variable	OR (95% CI)	p-value
Increasing age	1.10 (1.01, 1.21)	0.02
Higher pain	0.88 (0.78, 1.0)	0.03
Slower 40m SPW	1.07 (1.01, 1.14)	0.03
Increased hip flexor tightness	1.09 (1.01, 1.18)	0.03
Greater hamstring strength	21.67 (0.88, 533.30)	0.04
Previous smoker	0.23 (0.05, 1.2)	0.06
Increased hip flexion ROM	1.04 (0.99, 1.09)	0.11
Slower Timed Up and Go	1.16 (0.98, 1.38)	0.11
Poor pain beliefs/self efficacy	0.96 (0.91, 1.01)	0.12
Positive patellar compression test	0.31 (0.06, 1.59)	0.13
Greater quadriceps strength	4.56 (0.56, 36.96)	0.14
Greater hip abductor strength	47.15 (0.18, 12443.2)	0.15
20cm step test (< 50)	2.75 (0.64, 11.83)	0.16
Hip internal rotator strength	33.43 (0.10, 1158.39)	0.19
OR = odds ratio; 95% CI = 95% confidence interval; SPW = self-paced walk; ROM = range of motion		

Nine participants (11%) exhibited poor outcome at one year following physiotherapy intervention. This low number of events resulted in a wide 95% CI, suggesting findings are imprecise. Additionally nine participants and fourteen variables failed to meet the *a priori* decision to perform regression only when there were five times as many subjects as variables. Further analysis of poor outcome was considered inappropriate.

6.5.7. Sensitivity Analyses

Sensitivity analyses were performed to determine the degree to which decisions made about baseline variables or methods of analysis influenced results, and how changing parameters might affect the final models. Sensitivity analyses fulfil a number of functions including testing the robustness of the models, or possibly highlighting less than optimal solutions (Pannell, 1997).

Investigation of age and BMI as predictors of success following physiotherapy intervention: effect of cut-points on association with outcomes

Sensitivity analyses were performed on age and BMI data to assess the impact on multivariate models of treating these variables as either continuous data, or data with cutpoints. This was particularly important for age and BMI, as they are potential confounders of findings in studies of knee OA. Decisions regarding identification of cut-points in continuous data were outlined in Section 6.4.2.

For age, a cut-point of fifty-seven years was identified and univariate logistic regression was performed to investigate any association with successful outcome at one year. This univariate analysis was repeated on age after it was categorised into age bands. Age < 60 years was used as the referent category. Results are shown in Table 6.22 with data in continuous form for comparison.

Sensitivity analysis of BMI used categorical data only. The WHO classification for BMI was used to create categorical data, as outlined in Section 6.4.2 (World Health Organisation, 2006). No participants were in the underweight category. Univariate logistic regression was performed with normal weight as the referent category. Results are shown in Table 6.22 with data in continuous form for comparison.

Table 6.22 Univariate association of BMI and age with success at one year, displaying different data formats

Baseline variable/ Data format	OR (95% CI)	Significance
Age (continuous)	0.99 (0.95, 1.05)	p < 0.84
Age (categorical)		
< 60 years	Reference category	
60 years – 69 years	2.57 (0.70, 9.39)	p < 0.15
70 years – 80 years	1.33 (0.33, 5.43)	p < 0.69
Age (binary)		
> 57 years	4.94 (0.59, 41.53)	p < 0.08
BMI (continuous)	1.01 (0.95, 1.08)	p < 0.74
BMI (WHO classification)		
18.5 - 24.99 kg/m ² (normal)	Reference category	
≥ 25.0 - 29.99 kg/m ² (overweight)	1.7 (0.45, 6.48)	p < 0.44
≥ 30.0 kg/m ² (obese)	1.55 (0.41, 5.88)	p < 0.52
OR = odds ratio; 95%CI = 95% confidence interval; BMI = body mass index; WHO = World Health Organisation		

For age group categories, only the 60 – 69 years category showed a univariate association with success at one year (p < 0.15). Therefore multivariate analysis stratified by age was not performed. Dichotomised age (> 57 years) also showed a significant univariate association with success (p < 0.08). It was therefore retained for multivariate analysis with the resultant model shown in Table 6.23.

Regardless of data format, BMI did not exhibit univariate association with successful outcome at one year, indicating it does not predict response to

physiotherapy intervention. Changing decisions regarding format of BMI data did not alter multivariate models.

Table 6.23 Predictors of success at one year: multivariate sensitivity analysis for age (> 57 years)

Baseline Variable	Odds ratio (95% CI)	p-value
Disturbed sleep	3.92 (1.28, 11.95)	p < 0.016
Previous knee injury *	0.28 (0.09, 0.87)	p < 0.028
Posterior knee pain	3.15 (1.09, 9.06)	p < 0.033
p < 0.0008; AIC = 99.77; BIC = 109.45; GOF p = 0.41; AUC = 0.74		
95% CI = 95% confidence interval; * variable retained as <i>absence</i> of previous knee injury; AIC = Akaike information criterion; BIC = Bayesian information criterion; GOF = Hosmer Lemeshow goodness of fit; AUC = Area under the curve		

Accuracy statistics were calculated with the results given in Table 6.24. The optimal model that predicted success when age > 57 years was included was at least two out of three predictors, which gave a post-test probability of 60%. This test condition was met by 55% of participants. When less than two out of three predictors were present post-test probability dropped to 13%. This sensitivity analysis indicates that the original multivariate model is affected by decisions made about age.

Table 6.24 Post-test probability of success: sensitivity analysis for age (> 57 years)

	Number of patients in each group (%)*	Sensitivity (95% CI)	Specificity (95% CI)	LR+ (95% CI)	LR- (95% CI)	Post-test probability of success with +’ve test	Post-test probability of success with -’ve test
1 of 3	31 (37%)	100% (88.4, 100)	13.2% (5.5, 25.3)	1.14 (1.02, 1.28)	0.12 (0.01, 1.96)	38%	5%
2 of 3	32 (39%)	80% (61.4, 92.3)	60.4% (46.0, 73.5)	1.98 (1.36, 2.89)	0.35 (0.17, 0.71)	60%	13%
3 of 3	13 (16%)	30% (14.7, 49.4)	90.7% (77.9, 97.4)	3.0 (1.08, 8.34)	0.77 (0.60, 1.0)	67%	28%
Based on pre-test probability of success = 35%; * 7 participants (8%) had 0 of 3 variables							
95% CI – 95% confidence interval; LR+ = likelihood ratio of positive test; LR- = likelihood ratio of negative test; +’ve = positive test; -’ve = negative test							

Sensitivity analysis for prediction of outcome at one year, excluding participants who underwent joint arthroplasty during the study period

In the primary analysis participants who had joint arthroplasty during the one year study period were considered as non-responders at one year due to the confounding effect of surgery on outcome (section 3.7.2). However, regarding all joint arthroplasty participants as non-success may have diluted the strength of findings. An alternative is to exclude these participants from the analysis. This was not done in the primary analysis because it would have violated the intention to treat (ITT) principle and has instead been reserved for a sensitivity analysis. Analyses were repeated, except where numbers were previously found to be insufficient i.e. poor outcome at one year and analysis of usual care group. Results are therefore presented for predictors of success at one year; predictors of success at one year stratified by sex; and success at nine weeks as a predictor of success at one year, including stratification by sex.

Seventeen participants (15%) from the whole study sample underwent joint arthroplasty. Four of these were in the usual care group and thirteen in the treatment groups. Exclusion of arthroplasty participants produced a modified treatment cohort of seventy-two participants. Of these, thirty reached criteria for success at one year giving a pre-test probability of success of 42%.

Univariate logistic regression identified fifteen variables associated with success at one year ($p < 0.2$). Sex and instability were not significantly associated with outcome, unlike the primary analysis. Furthermore, greater hip internal rotator muscle strength and pain with knee extension were two different variables associated with success at one year (Table 6.25).

Table 6.25 Baseline predictors associated with success at one year in the modified cohort: univariate analysis

Baseline variable	OR (95% CI)	p-value
Disturbed sleep	5.33 (1.81, 15.76)	0.001
Higher pain	1.12 (1.04, 1.22)	0.003
Greater hamstring strength	11.16 (1.46, 85.18)	0.01
Presence of posterior knee pain	3.66 (1.33, 10.03)	0.01
Positive patella compression test	3.0 (1.13, 7.93)	0.02
Higher irritability	1.19 (1.02, 1.4)	0.02
Poor pain beliefs/self efficacy	1.05 (1.01, 1.09)	0.02
Knee locking	2.86 (1.07, 7.62)	0.03
20 cm step test (< 50)	2.61 (0.99, 6.87)	0.05
Greater hip external rotator strength	19.01 (0.49, 742.43)	0.10
Symptom duration (> 5 years)	2.21 (0.85, 5.74)	0.10
Pain with knee extension	2.22 (0.74, 6.64)	0.14
Previous knee injury	0.48 (0.17, 1.36)	0.16
Greater hip internal rotator strength	8.61 (0.39, 190.6)	0.16
Greater hip flexor strength	3.97 (0.55, 28.68)	0.16
OR = odds ratio; 95%CI = 95% confidence interval		

Multivariate logistic regression with backwards elimination produced the following model (Table 6.26). There was no collinearity between retained variables. Greater hip external rotator muscle strength was marginal for inclusion in the final model. However, when removed the model was weaker with a higher p-value and higher AIC. Greater hip external rotator muscle strength was therefore retained in the final model. As muscle strength was recorded as continuous data, it was not possible to generate clusters of variables, perform accuracy statistics or calculate post-test probabilities, as per the previous analyses.

Table 6.26 Baseline predictors of successful outcome at one year in modified cohort: multivariate sensitivity analysis

Baseline Variable	Participants with predictor (%)	Odds ratio (95% CI)	p-value
Disturbed sleep	42/72 (58%)	6.01 (1.75, 20.68)	p < 0.004
Posterior knee pain	26/72 (36%)	4.27 (1.24, 14.73)	p < 0.022
Previous knee injury	23/71 (32%)*	0.22 (0.06, 0.84)	p < 0.027
Greater hip external rotator strength†		0.03 (0.00, 1.90)	p < 0.098
p < 0.0002; AIC = 84.31; BIC = 95.62; GOF p = 0.43; AUC = 0.81			
95% CI = 95% confidence interval; * Missing data 1 participant; †Continuous variable; AIC = Akaike information criterion; BIC = Bayesian information criterion; GOF = Hosmer Lemeshow goodness of fit; AUC = area under the curve			

Multivariate analysis stratified by sex produced the models shown in Table 6.27. There was no collinearity between retained variables.

Table 6.27 Baseline predictors of successful outcome at one year in modified cohort: multivariate sensitivity analysis stratified by sex

Baseline Variable	Odds ratio (95% CI)	p - value
Male*		
Posterior knee pain	15.43 (2.03, 117.26)	p < 0.008
Higher irritability	1.44 (0.98, 2.12)	p < 0.064
Female†		
Disturbed sleep	21.48 (3.22, 143.10)	p < 0.002
Previous knee injury	0.11 (0.02, 0.74)	p < 0.023
*p < 0.0015; AIC = 33.37; BIC = 37.67; GOF p = 0.29; AUC =0.71		
†p < 0.0001; AIC = 43.73; BIC = 48.79; GOF p = 0.64; AUC = 0.73		
AIC = Akaike information criterion; BIC = Bayesian information criterion; GOF = Hosmer Lemeshow goodness of fit; AUC = area under the curve		

Irritability recorded using the i-score was not a dichotomous variable, so it was not possible to generate clusters of variables, perform accuracy statistics or calculate post-test probabilities, for males.

However, both retained variables for females are dichotomous, allowing analysis to be completed. Accuracy statistics and post-test probabilities were calculated, with the optimal models being 2 out of 2 predictors present giving a post-test probability of 80% (Table 6.28). This test condition was met by 35% of female participants. Failure to meet the test condition resulted in post test probability of success of 23%. (Note: pre-test probability of success in treatment cohort with joint arthroplasty participants excluded is 42%).

Table 6.28 Post-test probability for females: sensitivity analysis with modified cohort

	Number of participants in each group* (%)	Sensitivity (95% CI)	Specificity (95% CI)	LR+ (95% CI)	LR- (95% CI)	Post-test probability of success with +’ve test	Post-test probability of success with -’ve test
1 of 2	21/40 (53%)	100% (82.4, 100)	23.8% (8.2, 47.2)	1.3 (1.01, 1.67)	0.1 (0.01, 1.7)	48%	7%
2 of 2	14/40 (35%)	63.2% (38.4, 83.7)	90.5% (69.6, 98.8)	5.5 (1.63, 18.5)	0.42 (0.24, 0.76)	80%	23%
Based on pre-test probability of success = 42% (30/72); * 5/40 females (12%) had 0 of 2 predictors							
95% CI – 95% confidence interval; LR+ = likelihood ratio of positive test; LR- = likelihood ratio of negative test; +’ve = positive test; -’ve = negative test							

At nine weeks, threshold for success in this modified cohort was reached by thirty-seven participants (51%). The odds of success at one year given a successful response at nine weeks are shown in Table 6.29, including results stratified by sex. Results from the primary analysis are also shown for comparison, to highlight the effect of excluding joint arthroplasty participants.

Table 6.29 Odds of success at one year following success at nine weeks in the modified cohort: sensitivity analyses

	Odds ratio (95% CI)	p - value
Modified cohort (n = 72)*		
Nine week success	4.95 (1.77, 13.81)	p < 0.002
Nine week success (male)	13.33 (1.43, 123.99)	p < 0.023
Nine week success (female)	4.29 (1.14, 16.18)	p < 0.032
Primary cohort (n = 85)†		
Nine week success	4.81 (1.81, 12.8)	p < 0.002
Nine week success (male)	14.17 (1.59, 125.87)	p < 0.017
Nine week success (female)	3.86 (1.11, 13.46)	p < 0.034
*AUC = 0.69; †AUC = 0.68		
95% CI = 95% confidence interval; AUC = area under the curve		

Accuracy statistics and post-test probability for nine week response being a predictor of success at one year are shown in Table 6.30. Success at nine weeks increased the likelihood of success at one year from 42% to 60%. Failure to succeed at nine weeks decreases the likelihood of success at one year to 23%. Results were similar when stratified by sex: in males, failure to succeed at nine weeks decreased post-test probability of success to 10%; although this result lacked significance (CI crossed 1.0).

Table 6.30 Nine week response as predictor of success at one year in modified cohort: sensitivity analysis

	Number of participants in each group (%)	Sensitivity (95% CI)	Specificity (95% CI)	LR+ (95% CI)	LR- (95% CI)	Post-test probability of success with +’ve test	Post-test probability of success with -’ve test
Nine week response	72 (100%)	73.3% (54.1, 87.7)	64.3% (48.0, 78.4)	2.05 (1.30, 3.25)	0.42 (0.22, 0.78)	60%	23%
Nine week response (males)	32 (44%)	90.9% (58.7, 99.8)	57.1% (34.0, 78.2)	2.12 (1.25, 3.60)	0.16 (0.02, 1.07)	60%	10%
Nine week response (females)	40 (56%)	63.2% (38.4, 83.7)	71.4 % (47.8, 88.7)	2.21 (1.04, 4.72)	0.52 (0.27, 0.99)	61%	27%
Based on pre-test probability of success = 42%							
95% CI – 95% confidence interval; LR+ = likelihood ratio of positive test; LR- = likelihood ratio of negative test; +’ve = positive test; -’ve = negative test							

Summary of sensitivity analyses

BMI had no association with one year outcome regardless of data format. When age was dichotomised (> 57 years) the model predicting success also changed, to a model with only three variables and a lower post-test probability of success (60% compared with 66% for the original model). This suggests the model is sensitive to decisions about age data.

When participants who had undergone joint arthroplasty were excluded resultant models changed. While some variables were retained others were dropped and some new ones added. In the analysis of baseline predictors, post-test probability of success could only be calculated for females, increasing from 58% in the primary analysis to 80% (for two out of two predictors) in the sensitivity analysis.

Similarly, for nine week outcome as a predictor of success at one year, excluding participants who had undergone joint arthroplasty surgery increased post-test probability of success from 52/54% to 60%. However, as pre-test probability of success had also increased from 35% to 42%, the relative change due to the sensitivity analysis was marginal.

A summary of the main results is shown in Table 6.31

Table 6.31 Summary of main results

Description of outcome. (Pre-test probability of success/poor outcome)	Univariate association ($p < 0.2$)	Multivariate model ($p < 0.1$)	Optimal Test clusters/ % (number) meeting rule	Post-test probability of success
Primary analysis				
Success at one year (35%)	15 variables	6 variable model: Previous knee injury (OR 0.22)* Posterior knee pain (OR 3.65) Disturbed sleep (OR 3.42) Sex (female: male) (OR 3.23) Instability (OR 3.27) Symptom duration > 5 years (OR 2.64)	≥ 4 out of 6 (LR+ 3.59) 30% (n = 25) < 3 out of 6 (LR- 0.24) 34% (n = 28)	66% 11%
Success at one year: stratified by sex (35%)		Male (n = 40): Posterior knee pain (OR 8.44) Female (n = 45): Previous knee injury (OR 0.14)* Instability (OR 6.21) Disturbed sleep (OR 17.6)	Present (LR+ 3.01) 38% (n = 15) Absent (LR- 0.36) 62% (n = 25) ≥ 2 out of 3 (LR+ 2.56) 56% (n = 25) < 2 out of 3 (LR- 0.20) 44% (n = 20)	62% 16% 58% 10%

Description of outcome. (Pre-test probability of success/poor outcome)	Univariate association (p < 0.2)	Multivariate model (p < 0.1)	Optimal Test clusters/ % (number) meeting rule	Post-test probability of success
Success at one year: with specific sets of named predictors (35%)	4 out of 6 (weakest): female, disturbed sleep, posterior knee pain, symptom duration (> 5 years)			55%
	4 out of 6 (best): female, disturbed sleep, posterior knee pain, no previous knee injury			68%
	5 out of 6 (best): female, disturbed sleep, posterior knee pain, no previous knee injury, instability			87%
Secondary analyses				
Usual care group (18%)	15 variables 28 participants	Limited analysis due to insufficient numbers	Primary model did not fit the usual care data	Model did not fit the data
Nine week response (35%)	OR 4.81	-	Present (LR+ 2.02) Absent (LR- 0.42)	52% 18%
Nine week response: stratified by sex (35%)	Male OR 14.17	-	Present (LR+ 2.2) Absent (LR- 0.16)	54% 8%
	Female OR 3.86	-	Present (LR+ 2.05) Absent (LR- 0.53)	52% 22%
Nine week response plus baseline predictors (35%)		7 variable model (see previous ORs)	≥ 4 out of 7 (LR+ 2.34) 48% (n = 40) ≥ 5 out of 7 (LR+ 11.5) 22% (n = 18) < 3 out of 7 (LR- 0.13) 25% (n = 21)	56% 86% 7%

Description of outcome. (Pre-test probability of success/poor outcome)	Univariate association (p < 0.2)	Multivariate model (p < 0.1)	Optimal Test clusters/ % (number) meeting rule	Post-test probability of success
Poor outcome at one year (11%)	14 variables 9 participants	Unable to progress – insufficient numbers	-	-
Sensitivity Analyses: BMI				
Continuous data	p < 0.74	Did not progress	-	-
WHO categories (35%)	p < 0.44 – 0.52	Did not progress	-	-
Age				
Continuous data > 57 years (35%)	no association p < 0.08	Did not progress 3 variable model: Previous knee injury (OR 0.28)* Disturbed sleep (OR 3.92) Posterior knee pain (OR 3.15)	- ≥ 2 out of 3 (LR+ 1.98) 55% (n = 45) < 2 out of 3 (LR- 0.35) 45% (n = 38)	- - 60% 13%
Joint arthroplasty excluded (n = 72) (42%)	15 variables	4 variable model: Disturbed sleep (OR 6.01) Posterior knee pain (OR 4.27) Previous knee injury (OR 0.22)* Hip external rotator strength (OR 0.03)	Unable to progress	-

Description of outcome. (Pre-test probability of success/poor outcome)	Univariate association (p < 0.2)	Multivariate model (p < 0.1)	Optimal Test clusters/ % (number) meeting rule	Post-test probability of success
Joint arthroplasty excluded (n = 72): stratified by sex (42%)		Male (n = 32): Posterior knee pain (OR 15.43) Irritability (1.44) Female (n = 40): Disturbed sleep (OR 21.48) Previous knee injury (0.11)*	Unable to progress 2 out of 2 (LR+ 5.5) 35% (n = 14) < 2 out of 2 (LR- 0.42) 65% (n = 26)	- 80% 23%
Joint arthroplasty excluded (n = 72) Nine week success: Stratified by sex (42%)	OR 4.81 Male OR 14.17 Female OR 3.86		Present (LR+ 2.05) Absent (LR- 0.42) Present (LR+ 2.12) Absent (LR- 0.16) Present (LR+ 2.21) Absent (LR- 0.52)	60% 23% 60% 10%‡ 61% 27%
<p>OR = odds ratio; * variable retained as <i>absence</i> of knee injury; LR+ = likelihood ratio with positive test; LR- = likelihood ratio with negative test; BMI = body mass index; WHO = World Health Organisation; †continuous variable; ‡non-significant result Note: 95% CI included in text and chapter tables</p>				

7. Study to identify predictors of outcome: discussion

7.1. Chapter overview

The primary aim of the thesis is to identify baseline predictors of outcome following physiotherapy intervention for patients with knee osteoarthritis (OA).

The specific research questions addressed in the prognostic study were:

- What variables assessed at baseline predict success at one year for participants with knee OA following physiotherapy intervention?
- Do identified variables predict treatment response or natural history of knee OA disease over one year?
- Does the immediate post-treatment response at nine weeks predict outcome at one year follow-up?
- What variables predict poor outcome at one year for participants with knee OA following physiotherapy intervention?

The extent to which the results from the prognostic study have answered the research questions are discussed in this chapter. Predictive variables are discussed with reference to previous evidence of their role in knee OA, possible underlying mechanisms, and clinical implications. Other variables previously shown to have a role in disease progression, but not identified as predictors following physiotherapy intervention, will also be discussed. The strengths and limitations of the prognostic study will be outlined, followed by discussion of clinical and research implications.

7.2. Predicting success with physiotherapy for patients with knee OA

Six baseline variables were identified as predictors of successful outcome at one year. These were posterior knee pain, knee pain causing disturbed sleep, absence of previous knee injury, instability, female sex, and symptom duration (>

5 years). When a participant with knee OA had at least four out of six of these predictors at baseline assessment, likelihood of successful outcome with physiotherapy almost doubled, with probability of success increasing from 35% to 66%. When participants had two or less predictors, likelihood of success dropped to 11%. These noticeable changes in probability provide useful information to help health practitioners make clinical decisions and treatment choices. Additionally, patients with knee OA should be better informed regarding likely outcomes from physiotherapy treatment, helping them to make an informed choice regarding an effective alternative to joint surgery.

A new approach to this type of study may enable health practitioners to further improve prediction of outcome for an individual patient by considering the exact make-up of the model, i.e. the specific predictors present, rather than just the optimal number of predictors. For example, in a four out of six model (*absence* of previous knee injury, female sex, posterior knee pain, and disturbed sleep), substitution of absence of previous knee injury for symptom duration (> 5 years), decreased probability of success from 68% to 55% (Table 6.11). Differences such as this in probability of success could feasibly influence clinical decision-making for individual patients. This level of detail has not previously been examined in clinical prediction rules (CPR), and is a useful adjunct to analysis.

All models complied with *a priori* levels of significance. They were also able to discriminate between responders and non-responders as demonstrated by area under the curve (AUC) values (range 0.68 to 0.91). Hosmer Lemeshow tests confirmed that the models fit the data.

7.2.1. Posterior knee pain

Posterior knee pain was an unexpected finding as a predictor of outcome, although pain is a common feature of knee OA, and a common cause for consultation in primary care (Bedson et al., 2007). It was identified as a predictor

following the primary analysis; the primary analysis stratified by sex; and was retained following sensitivity analyses (Table 6.30). This suggests it is a robust finding with a strong association with successful outcome.

The most common sites of pain in knee OA are generalised or medial knee pain (Wood et al., 2007; Thompson et al., 2009). In the current study 36% of participants in the treatment cohort reported posterior knee pain; in all cases this was in conjunction with other sites of pain. Occurrence of isolated posterior knee pain was 0% which compares with 5% reported in a previous study (Thompson et al., 2009). Different methods of recording pain location limit direct comparison, a difficulty which could be overcome by standardising recording of pain patterns and location, as described in recent publications (Wood et al., 2007; Thompson et al., 2009). These methods could be used to determine if isolated posterior knee pain or generalised knee pain is a predictor of treatment success for patients with knee OA, as well as establishing reliability of pain report.

It is feasible that posterior knee pain is elicited by stressing a tight or inflamed posterior joint capsule and supporting ligaments. The role of these periarticular tissues as a source of symptoms in knee OA is increasingly recognised (Moskowitz et al., 2007). Decreased knee extension, fixed flexion deformity and increased hamstring activity are common clinical features of knee OA (Hortobagyi et al., 2005; Reid and McNair, 2010). Additionally there is discomfort associated with stretching tight muscles (Reid and McNair, 2010). Knee pain can also be referred from proximal structures in the hip or lumbar spine. Participants received interventions that addressed these impairments, making posterior knee pain a modifiable baseline variable. This may have contributed to it being identified as a predictor of treatment success.

Clinical Implications

Identification of posterior knee pain as a predictor of success could influence practitioners to consider pain location when assessing patients with knee OA.

While routine clinical practice for physiotherapists is to chart pain location, confirmation of presence or absence of posterior knee pain is probably less common. Further investigation may clarify whether pain report itself is sufficient to establish pain location, or whether physical testing to reproduce pain is required.

In general, posterior knee pain was a robust predictor that appeared in different versions of the prognostic models with significant and strong associations with outcome (range of multivariate odds ratios (OR) 3.15 to 15.43). There was also a strong association with males that remains unexplained.

7.2.2. Absence of previous knee injury

A recent systematic review identified previous knee injury as an important risk factor for the development of knee OA (Blagojevic et al., 2010), although its role in progression was less clear. The systematic review (Chapter 2) found limited strength of evidence, based on one study (Cooper et al., 2000), that previous knee injury was *not* a predictor of progression. Overall, evidence suggests that risk factors for development of disease may differ from those for progression. However, as previous knee injury appears to have an influence on both, it was included as a potential baseline predictor of outcome following physiotherapy intervention.

In this prognostic study 35% of participants recalled previous knee injury requiring treatment, which is similar to 28% of participants in a previous study who reported being unable to walk unaided for one-week (Cooper et al., 2000). Analysis revealed previous knee injury had a low OR (0.22), meaning it was very unlikely to predict success following physiotherapy intervention. An alternative method of reporting is to state *absence* of knee injury predicts success, which was more useful for subsequent analysis.

In the current study assessment of previous knee injury did not include any classification of injury type or time elapsed since injury. Therefore it was not possible to determine whether different types of injury were associated with differing odds of outcome. Similarly, it could not be determined whether old injuries or current injuries were more likely to influence odds of success. Physiotherapy interventions in the clinical trial were designed to target the clinical features of knee OA. Previous knee injury may have produced different limitations (longstanding adaptations to joints from old injuries, or possibly more acute reactions from recent injury), which were not addressed by the knee OA specific treatment, and therefore unlikely to result in successful outcome. Accordingly it seems reasonable that *absence* of knee injury would be associated with successful outcome. It was included in all the prognostic models including those following sensitivity analyses, suggesting it is a robust variable with a strong association with successful outcome following physiotherapy intervention (Table 6.30).

Future research should involve more detailed investigation of injury types, including dates and previous treatment, to determine the features most relevant for prediction of treatment response.

Clinical Implications

Study findings suggest it is important to question patients about previous knee injuries, as information gained will help predict likely outcome with or without physiotherapy intervention. Residual impairments from previous knee injury should be carefully evaluated and guide selection of treatment techniques, as the treatment protocols designed for knee OA patients may not adequately address these impairments.

7.2.3. Disturbed sleep

In the prognostic study, 60% of participants reported sleep disturbance due to the pain of their knee OA. Sleep disturbance is a common feature of rheumatologic disease (Abad et al., 2008). In the OASIS study of 429 older individuals with knee pain or knee OA, many reported problems at least weekly with sleep onset (31%), sleep maintenance (81%), or early morning waking (51%) (Wilcox et al., 2000). A more recent study reported 81% of participants with hip or knee OA experienced night pain, that could occur at different stages of disease, and was not directly related to disease severity (Woolhead et al., 2010). The complexity of the issue, and associated implications for assessment (Woolhead et al., 2010), may have contributed to the lower reported sleep disturbance in the current study; specifically in light of the decision to dichotomize the variable.

Future research could address these issues by investigating the different features of sleep disturbance including frequency of waking, duration of waking, and remedial action required to get back to sleep. This would give a more comprehensive assessment of the problem and possibly identify the characteristics most relevant for using disturbed sleep as a predictor of outcome.

Disturbed sleep was identified as a predictor of successful outcome in the primary analysis; for females in the analysis stratified by sex; and in all the sensitivity analyses. These findings suggest disturbed sleep is a robust variable for predicting successful outcome following physiotherapy treatment. It was potentially modifiable by trial interventions that addressed pain and other impairments such as muscle weakness, soft tissue restriction, and joint stiffness (Abbott et al., 2009). Treatment may have increased functional capacity while limiting exacerbation of symptoms, contributing to better sleep and successful outcome. However it is also possible that presence or absence of night pain reflects the fluctuating course of the disease, including flare-ups of inflammation,

that are independent of treatment effects of physiotherapy. Additionally the presence of pain disturbing sleep may be sufficient to prompt an individual to seek to treatment, making it a common clinical feature (60% of the treatment group reported disturbed sleep – 51/85 participants).

Clinical implications

Night pain or disturbed sleep is used frequently as an indicator of the need for joint arthroplasty (Woolhead et al., 2010). However, findings from the current study show that participants experiencing disturbed sleep due to the pain of their knee OA could achieve a successful outcome with physiotherapy. The complexity of the issue has been highlighted, suggesting clinical assessment should include detailed questioning. Information could be used to identify and monitor treatment goals, in addition to predicting outcome.

7.2.4. Instability

As indicated in Chapter 5, functional knee instability is defined as a sensation of buckling, slippage, or giving way (Fitzgerald et al., 2004a). It is a common clinical feature in a wide range of disorders including knee OA, and is included in many patient self-report instruments (Irrgang et al., 1998). Previous work has demonstrated an association between instability and greater functional limitation (Fitzgerald et al., 2004a; Schmitt et al., 2008).

In the current study 29% of participants reported knee instability, with 27% stating it affected their everyday activities. In comparison, a previous study of 105 subjects with knee OA reported 63% of participants with instability, and for 44% this instability affected everyday activities (Fitzgerald et al., 2004a). Inclusion criteria for the current study were based on American College of Rheumatology (ACR) clinical criteria (Altman et al., 1986) and did not include any radiological examination. It is possible this resulted in the inclusion of more individuals with mild disease and little functional impairment compared to the previous study.

Additionally, current results indicate an association of instability with female (but not male) participants (Table 6.8). A higher percentage of females in the earlier study (80% compared to 56% here) could have contributed to higher reported instability (Fitzgerald et al., 2004a).

Evidence demonstrating effectiveness of treatment for instability is scarce, and there is a need for research to develop and test such treatments (Fitzgerald et al., 2002; Bennell and Hinman, 2005). A recommendation to include interventions to address instability was adopted in the MOA clinical trial (Fitzgerald et al., 2004a; Abbott et al., 2009), which may have contributed to the identification of instability as a predictor of success.

Instability was identified as a predictor in a limited number of the prognostic models. It was not retained following sensitivity analyses, suggesting it is a less robust predictor of outcome than posterior knee pain, disturbed sleep or absence of knee injury (Table 6.30).

Clinical implications

Patients presenting with knee OA should be questioned about knee instability. It is a feature that can potentially be modified by physiotherapy treatment, including strength, balance and coordination exercises, all of which have crossover benefits for other impairments associated with knee OA.

7.2.5. Sex

Although sex is not a modifiable factor, it can be a useful predictor of treatment outcome for patients with knee OA. Women with knee OA had greater odds of being successful with physiotherapy treatment than men (OR 3.23), except when men had posterior knee pain (OR 8.44) (Table 6.30). In some analyses there was a marked increase in probability of success for women; for example, in the stratified analysis of the “joint replacement excluded” cohort, there was a strong two-variable model for females. In this sample, women with disturbed

sleep and absence of previous knee injury had an 80% probability of success following physiotherapy treatment (Table 6.27). This is a noteworthy increase, further strengthened by 35% of female participants meeting the test conditions.

The new post-estimation analysis revealed that when the same variables were present, women had a greater probability of success than men. For example women with posterior knee pain, disturbed sleep, instability, and absence of knee injury had a probability of success of 87%, compared to 68% probability in men with the same four variables (Table 6.11). However, when sensitivity and specificity were taken into account in the more conventional analysis, post-test probability for the stratified primary analysis was very similar for women (58%) and men (62%) (Table 6.30).

A previous study described female sex as a “stable” predictor which predicted beneficial outcome following rehabilitation for knee and hip OA with each of three outcome measures (Weigl et al., 2006). In contrast, sex was not retained in all models following sensitivity analyses in the current study, suggesting models were susceptible to change, and questioning the stability of female sex as a predictor of successful treatment response.

There is ample evidence that women are more likely to develop knee OA than men (Srikanth et al., 2005; Blagojevic et al., 2010). Conversely, there is conflicting evidence for sex as a predictor of natural progression (Chapter 2). Women also report higher pain intensity for the same radiographic findings (Srikanth et al., 2005; Cho et al., 2010), poorer quality of life (Debi et al., 2009), as well as more disability, depression, and catastrophizing (Keefe et al., 2000). For females the accumulated evidence implies that they are likely to have more severe symptoms, as well as less effective coping strategies. Therefore the increased probability of success for female participants is an important finding. It is feasible that the comprehensive rehabilitation programme addressed not only

physical impairments, but may also have induced psychological benefit in terms of improved understanding of knee OA, provided reassurance, reinforced beneficial behaviours, and increased confidence in functional performance.

Clarification of the role of sex as a predictor should be the subject of future research. Additionally, analysis stratified by sex should be considered in studies involving patients with knee OA.

Clinical implications

Health practitioners need to recognise that women with knee OA may report more severe symptoms than men and display more psychological distress. However, these features do not preclude success with physiotherapy intervention, and these early results indicate that women may have a greater probability of better outcomes at one year.

7.2.6. Symptom duration (> 5 years)

Multivariate analysis showed that longer symptom duration (> 5 years) was a predictor of successful outcome (OR 2.64), and is one of the six variables in the main prognostic model. In contrast, previous studies have associated longer duration of symptoms with poor clinical outcome in a range of musculoskeletal conditions, including knee pain (van der Waal et al., 2005; Mallen et al., 2007b; Jinks et al., 2008).

It seems counterintuitive that longer duration of symptoms is a predictor of treatment success, though there is only limited evidence that longer symptom duration is associated with radiographic progression of knee OA (Wolfe et al., 2002). There is no evidence to suggest it is equivalent to greater severity of disease. It is not known whether physiotherapy intervention is more effective at specific stages of disease, and this cannot be concluded from the current study as there was no classification of stage or severity of disease incorporated into the design. However, it is plausible that it would be of greater benefit in early or mid

stage disease, while there is better opportunity to modify structures such as muscle, ligament, or capsule.

An alternative perspective is that by five years participants may well be more accepting and less fearful of their diagnosis, have gained understanding of how to self-manage their condition, and be knowledgeable regarding effective medication. Another consideration is the slow progression and variable course of knee OA, with a third of patients experiencing deterioration, while a third stay the same and a third improve (National Collaborating Centre for Chronic Conditions, 2008; Dekker et al., 2009; Bijlsma et al., 2011). All these factors could provide a suitable window of opportunity for effective physiotherapy intervention, after disease has become established but before it becomes too severe, which may contribute to symptom duration (> 5 years) being a predictor of successful treatment response.

It should be noted that longer symptom duration (> 5 years) was borderline for inclusion in the multivariate model ($p = 0.096$). A decision to retain it in the model was made on the basis of inferior statistical performance when it was removed. However, it was not retained as a predictor in stratified analyses or sensitivity analyses suggesting it was not a robust finding. In the current study recording of symptom duration relied on participant recall. It is recognised that patients have difficulty recalling previous health status (Guyatt et al., 2002; Kamper, 2009), and therefore recall of disease duration, where knee pain is often the first symptom, is likely to be unreliable (Spector et al., 1994).

Future research should ensure as much detail as possible is obtained about symptom duration as a predictor of treatment response. This unexpected finding requires replication in future studies. However, if confirmed it will provide valuable information, and corroborate other research that challenges patient and health practitioner beliefs that diagnosis of knee OA leads to an inevitable decline

and probable joint surgery (National Collaborating Centre for Chronic Conditions, 2008; Hunter, 2009).

Clinical implications

Health practitioners can reassure patients with knee OA that progression of disease is not inevitable and when it does occur it may be a slow process taking place over years. Longer duration of symptoms does not necessarily result in more severe disease and the current findings suggest that if symptoms have been present for more than five years there is a greater likelihood of success with physiotherapy intervention. However, clinicians should treat these results with caution as they have borderline significance and are sensitive to change.

7.2.7. Hip external rotator muscle strength

Muscle strength of hip external rotators emerged as a predictor of success in the sensitivity analysis with the modified cohort (exclusion of participants who underwent joint arthroplasty). All muscle strength variables were recorded using continuous data, with higher values representing a better condition. Therefore the low OR (0.03) indicates that strong hip external rotators were (highly) unlikely to predict success. For ease of understanding, this finding can be re-interpreted as weak hip external rotators predicted success.

A recent study demonstrated significant difference in strength of all hip muscle groups between subjects with knee OA and asymptomatic individuals, with the greatest difference being for hip external rotators (Hinman et al., 2010). Another study from the same group of researchers showed that including hip muscle strengthening in a rehabilitation programme for knee OA participants improved pain and function (Bennell et al., 2010). These findings provide a plausible explanation for hip external rotator weakness being a predictor of success, as it is a modifiable factor targeted by the clinical trial secondary interventions. Physiotherapists providing the interventions were encouraged to

incorporate additional techniques, based on impairment criteria and reflecting usual clinical practice (Abbott et al., 2009). Further analysis of data could reveal how many knee OA participants actually performed hip strengthening exercises, which would support or refute this point. However, for the prognostic study, the investigator was blinded to group allocation and treatment received.

Clinical implications

Clinically these findings suggest that assessment and physiotherapy intervention for knee OA should consider impairments at the hip, and specifically there is some evidence that weakness of external rotator muscles can be used to predict successful outcome at one year.

7.2.8. Irritability

Another finding from the sensitivity analysis of the modified cohort (exclusion of participants who underwent joint arthroplasty), was that for males, in addition to posterior knee pain, an increase in irritability score (the i-score) was a predictor of successful outcome. It should be noted that in the modified cohort, the i-score ranged from 17 – 32 (median 23.5) out of a maximum score of forty, indicating moderate levels of irritability for most participants. Treatment was individually tailored to participants, with vigour of technique based partly on informal assessment of irritability by physiotherapists providing the interventions (Abbott et al., 2009). This meant that, as with other predictors identified in this study (posterior knee pain, disturbed sleep, hip external rotator muscle weakness), irritability was a modifiable baseline variable addressed by the trial interventions. Consideration of irritability is likely to have contributed to good clinical outcomes by matching a patient's reported response to movement with selection of treatment technique at an appropriate level of vigour.

It is possible that higher levels of irritability may have caused participants to seek treatment during a more active phase of disease, as they were unable to self-

manage, exacerbating symptoms by underestimating their own sensitivity to movement and joint forces. There was therefore a greater potential for improvement.

Irritability as a component of symptom behaviour has not previously been investigated in patients with knee OA, nor has it been identified as a predictor of treatment response. These findings demonstrate the usefulness of a scale assessing irritability, as well as limited evidence for its role as a predictor of treatment response in patients with knee OA. Further research is necessary to confirm these findings and could explore the use of the i-score as a predictor of treatment response for other interventions for knee OA.

Clinical implications

The current study reinforces the importance of considering irritability when tailoring physiotherapy to individual patients. However, there is limited evidence to support using the i-score as a predictor of success, as it was only identified in a small sub-group of male patients in the modified cohort.

7.3. Variables not identified as predictors

There is strong evidence for BMI, age, and malalignment as predictors of knee OA progression (Chapter 2). Additionally, BMI and age may be confounders of results in clinical trials investigating interventions for knee OA, while BMI and malalignment are potentially modifiable features. Other modifiable features of knee OA that are commonly investigated are quadriceps muscle weakness and pain. However, none of these variables were identified as predictors of success in the primary analysis.

BMI did not exhibit univariate association with any of the outcomes despite being investigated thoroughly using different cut-points and analysis stratified by WHO classification. The conclusion is that increasing BMI is not a barrier to successful intervention with physiotherapy for patients with knee OA.

Sensitivity analysis using age as a dichotomised variable (> 57 years) showed a univariate association of age (> 57 years) with successful outcome (OR 4.94, $p < 0.08$), and a 3-variable model following multivariate analysis that differed from the primary 6-variable model. Identified predictors, posterior knee pain, disturbed sleep, and *absence* of previous knee injury, were in both models. The 3-variable model accounts for age and is simpler than the primary model, producing similar changes to post-test probability of success. There is an increase to 60% when two or more variables are present, and a decrease to 13% when less than two variables are present (Table 6.23). There is the added advantage that 55% of participants met the optimal test condition of two or more predictors, giving relatively narrow 95%CI and increasing confidence in the result. This finding demonstrates the primary model was sensitive to change, and that the new model performs in a similar manner. With fewer variables to remember, the simpler model may have greater clinical appeal.

Valgus or varus malalignment were not associated with any outcomes. The method of visual assessment of alignment, with only “fair” inter-rater reliability (Table 4.4) could have contributed to this finding, as some participants with malalignment may have been missed or incorrectly classified. Although malalignment is potentially modifiable using orthotics or bracing (Hinman et al., 2009; Foroughi et al., 2010), neither of these interventions were included in the current clinical trial.

Quadriceps muscle strength or weakness was not identified as a predictor in any of the models, with the exception that greater quadriceps strength had a univariate association with poor outcome (OR 4.56, $p < 0.14$). This finding lacks credibility and is likely to be due to the small number of poor outcomes producing unreliable results (Table 6.20). The lack of association with success is surprising as the variable is modifiable and the trial interventions included

quadriceps muscle strengthening. However, previous research has identified a complex interplay between malalignment, quadriceps strength, and joint laxity that may mediate treatment response (Sharma et al., 2003b; van der Esch et al., 2006; Lim et al., 2008; Hunter et al., 2009b). This could mean that none, in isolation, would be a good candidate as a predictor, and that local mechanical factors should be investigated together to assess their effect on treatment outcome or disease progression (Sharma et al., 2003a). Additionally, quadriceps muscle weakness may be influenced by impaired muscle activation, with conflicting reports of the effect this might have on treatment response, thereby complicating its role as a predictor (Fitzgerald et al., 2004b; Hortobagyi et al., 2005; Petterson et al., 2008; Scopaz et al., 2009).

These findings highlight the complexity of patient presentations, and indicate that analysis of factors in isolation may oversimplify the clinical situation. Although the multivariate analysis allows for interaction of variables, a preceding step in the analysis involves investigation of variables in isolation, which may exclude them or mask their potential impact. With increasing evidence from biomechanical studies and randomised controlled trials (RCT) it may be possible to develop composite variables that would perform better as predictors.

Pain did have a small but significant univariate association with success at one year (OR 1.06, $p < 0.07$), with poor outcome at one year (OR 0.88, $p < 0.03$), and with success in the modified cohort (exclusion of participants who underwent joint arthroplasty) (OR 1.12, $p < 0.003$). However, pain was not retained in any models following multivariate analysis, meaning it is not an independent predictor of outcome. ORs were all close to 1.0, meaning little difference in odds of outcome with pain present or absent.

The omission of pain as a predictor of outcome is perhaps not surprising given the variability in pain experience for patients with knee OA (Hawker et al., 2008b). In the early stages of disease it is more likely to be inconsistent or at least fluctuate in intensity in relation to mechanical loading or positioning, compared to more consistent pain in later stages (Neogi et al., 2010). This leads to difficulty in assessment of OA pain. In recognition of this and the inadequacy of existing measurement tools, a new pain measure was developed to address dimensions of pain considered important by patients with hip and knee OA (Hawker et al., 2008a). It includes questions on the intermittent or constant nature of pain, pain intensity, and impact on daily life and has good psychometric properties (Hawker et al., 2008a). Research is lacking about how pain changes with disease progression (Hawker et al., 2008b), and evidence is conflicting for its role as a predictor of such progression (Chapter 2). While pain is accepted as the most important symptom of knee OA and the main reason patients consult a health practitioner (Jordan et al., 2006; Bedson et al., 2007), it does not appear to have a role in the prediction of outcome following physiotherapy intervention.

Clinical implications

Results from this study suggest malalignment, body mass index (BMI), and quadriceps weakness are not predictors of success with physiotherapy intervention, while age and pain have a limited role. It should be encouraging for clinicians that neither BMI nor increasing age were barriers to success for participants in this study. However given the limitations and early stage of this research, caution should be exercised about applying this information in clinical practice.

Similarly, higher severity of pain did not predict outcome. However, pain assessment should include questions on the temporal nature of pain, as well as pain intensity, site of pain, and impact on function. A better understanding of the

patient's symptom behaviour and possible underlying mechanisms may help direct treatment selection (Hawker, 2009).

There are likely to be sub-groups of patients with knee OA who will respond differently to the same treatment. The complicated interplay of local mechanical factors and how they might predict outcome is unclear. The lack of a simple reliable method to assess malalignment clinically could have influenced the current findings regarding its role as a predictor of outcome following physiotherapy intervention. Extensive evidence of its importance in knee OA (Hunter et al., 2009; Sharma et al., 2010) means clinicians should not neglect to assess and manage knee malalignment when detected in their patients.

7.4. Secondary analyses

7.4.1. Do identified variables predict response to physiotherapy intervention or natural progression over one year?

Results from the systematic review (Chapter 2) identified several variables with strong evidence as predictors of knee OA natural progression: age, varus alignment, multiple joint OA, radiographic features, and BMI (for longer duration). Age, BMI, and varus alignment were examined as potential predictors in the current study and none were found to be associated with success following physiotherapy intervention. Furthermore, none of the variables identified in the prognostic models have strong evidence from the literature as predictors of natural progression. This would support the view there are different predictors of treatment response *versus* natural progression.

The ability of baseline variables to predict response to physiotherapy intervention, rather than predicting natural progression, is the basis of the second research question. The current study was nested within an RCT, and contained a

usual care group of twenty-eight participants. However, this number was insufficient to allow full analysis with multivariate logistic regression.

An alternative analysis was performed using the whole clinical trial study sample of participants with knee OA (n = 113), with a treatment variable forced into the model. The Hosmer Lemeshow goodness of fit (GOF) statistic was $p = 0.04$, indicating significant difference between observed and predicted values, meaning the model was not a good fit for the data (Table 6.12). No conclusion can be drawn from this result regarding the ability of identified variables to predict outcome at one year, with or without treatment.

In a different analysis, results showed that data from the usual care group, when applied to the primary six-variable model, did not produce any significant findings (Table 6.13). This provides limited evidence that the six-variable model is a predictor of treatment response rather than the natural course of knee OA.

Future research using RCT study design with a larger control group is necessary to allow complete multivariate analysis and identification of predictors of outcome in the absence of physiotherapy treatment. This would allow direct comparison of predictors of treatment response, and thus differentiate between non-specific predictors of natural progression and treatment effect modifiers (Stanton et al., 2010).

7.4.2. Nine week response as predictor of one year outcome

Another research question examined whether immediate post-treatment response at nine weeks predicted outcome at one year follow-up. Hypothetically, immediate post-treatment response could be a stronger predictor of outcome at one year than baseline variables.

Immediate post-treatment success at nine weeks compared with non-success at nine weeks increased the odds of a positive outcome at one year (OR 4.81), an effect accentuated in males (OR 14.17). However, incorporating sensitivity and

specificity into the analysis produced more modest likelihood ratios (LR) resulting in post-test probabilities of success at one year of 52% (all), 54% (male), and 52% (female); lack of success at nine weeks resulted in post-test probability of success at one year of 18% (all), 8% (male), 22% (female) (Table 6.16). Decreased probability of success for males (8%), while dramatic, lacked significance. These findings indicate nine week response is a weaker predictor of one year outcome than baseline variables.

Nine week successful response was combined with the six-variable baseline model to produce a seven-variable model, to investigate if probability of success could be further increased. This reflects the clinical situation where information is available for baseline variables and response to treatment can be evaluated. While the model did produce marked changes in probability (Table 6.17), this is only relevant if the optimal test conditions contained nine week response. The new approach, using a post-estimation test as described previously, was used to examine this (Table 6.18).

The test condition “at least 5 out of 7” predictors present produced an 86% post-test probability of success. Examining the probability of success for this condition when nine week success was one of the five predictors revealed a range of probabilities between 73% - 88% (Table 6.18). These results represented an increase from 66%, the post-test probability of success for the optimal version of the six-variable model. Comparing results of these analyses in the six-variable and seven-variable model demonstrates the value of adding nine week response to baseline predictors when calculating probability of success at one year. For example, women with posterior knee pain, absence of previous injury, disturbed sleep, and longer duration of symptoms (> 5 years) had an 85% probability of success (Table 6.11), while adding nine week response improved this probability even further to 93% (Table 6.18).

The ability to predict long-term outcome at different stages of clinical assessment or management, or to adapt prognosis over time is well suited to physiotherapy clinical practice, as patients are often seen on multiple occasions for a programme of care. Serial evaluation enriches the information available to clinicians, and its ability to determine long-term outcome at different time points was previously investigated in an acute low back pain population (Wand et al., 2009). Evidence from the current study highlights the value of the post-treatment response, combined with baseline variables, to prediction of long-term outcome for participants with knee OA.

Clinical implications

The purpose of identifying predictors of success at one year is to aid in decision-making at baseline. There are a number of reasons why health practitioners and patients may elect to trial a course of conservative management, including presence of co-morbidities, patient preference, and rationing of health-care (e.g. joint arthroplasty). Results of this analysis suggest that trialling physiotherapy intervention to see if there is any post-treatment benefit gives less indication of successful outcome at one year than using baseline predictors. As there is a cost involved in provision of physiotherapy treatment it would seem sensible to evaluate patients for likely success *prior* to referral or at baseline assessment.

However, if baseline assessment identifies at least four out of six predictors *and* the patient has a successful outcome at nine weeks, patients and health practitioners can be reassured of an even greater probability of continued success at one year. Also in the absence of adequate baseline assessment or compulsion to provide physiotherapy treatment, all stakeholders can be reassured that good response to treatment increases likelihood of success at one year.

7.4.3. Predictors of poor outcome at one year

An additional research question was to identify predictors of poor outcome following physiotherapy intervention. Knowledge that a patient is highly likely to get worse, in spite of physiotherapy intervention, would influence primary care clinicians to try alternative therapeutic management. An important point is that participants who worsened are a distinct sub-group of those classified as “not successful at one year” (i.e. didn’t meet OMERACT-OARSI responder criteria). They, therefore, needed separate criteria for defining response (Section 5.8.6).

In this study, the number of participants experiencing deterioration was insufficient to allow complete analysis, with only nine of the eighty-five participants who received physiotherapy intervention classified as having a poor outcome at one year. This finding is not unexpected as most clinical trials are designed to ascertain clinical improvement, and the numbers required to identify deterioration are much larger (Angst et al., 2002; Mannion et al., 2009). Thus analysis was restricted to identifying baseline variables with univariate *association* with poor outcome. Many of these variables were different from those associated with success. However, the low ORs, wide confidence intervals (CI), and occurrence of CI crossing 1.0, limit the usefulness of any interpretation. Future research with larger numbers is required to determine if different baseline variables are useful for predicting success or worsening.

7.5. Sensitivity analyses

Results of the sensitivity analyses are provided in Section 6.5.7 and have been discussed in preceding sections of this chapter. Dichotomising the age variable changed the prognostic model, as did the decision to exclude participants who had undergone joint arthroplasty (Table 6.29). However, the change in values of post-test probability was minimal and therefore added little

benefit over the original models. The exception to this was the sensitivity analysis, stratified by sex, of the modified cohort (excluding participants who had undergone joint arthroplasty). If women experienced sleep disturbance and had absence of knee injury, their probability of success increased from 42% to 80%; if they didn't exhibit these two characteristics probability of success dropped to 23%.

It may be argued that producing numerous models to encompass different scenarios obscures the main findings: it complicates the message and may provide a barrier to implementation. Conversely, failure to conduct a sensitivity analysis means the strength of the model has not been fully tested (Pannell, 1997). Additionally opportunities would be missed to simplify the model, or develop an improved version, with potential to predict success more accurately.

Change to the primary model following sensitivity analysis decreases confidence that this is the optimal model, and reinforces the need for further study. Posterior knee pain, disturbed sleep, and absence of knee injury were retained in different versions of the model, including following sensitivity analysis, suggesting they are robust and important predictors. Variables such as instability, sex, longer duration of symptoms, irritability, and hip rotator muscle strength were sensitive to decisions made during analysis. Nevertheless, all these variables should be included in future study of predictors of treatment response in patients with knee OA. Age should also be included in the analysis, as sensitivity analysis revealed a univariate association of the dichotomised age variable with success at one year, and subsequent inclusion in multivariate analysis changed the primary prognostic model. Researchers should consider carefully the inclusion of participants likely to undergo joint replacement, as decisions on how to handle data from these subjects clearly affected the results in the current study.

7.6. Comparison with previous findings

The current study has identified six main predictors of successful outcome following physiotherapy intervention, plus a few additional ones from the sensitivity analysis. With the exception of hip external rotator muscle weakness, all variables were demographic features (sex), self-reported symptoms (posterior knee pain, disturbed sleep, instability, i-score) or items related to disease history (absence of previous injury, symptom duration). This is similar to many other studies of predictors of success following physiotherapy intervention (Stanton et al., 2010). There are relatively few clinical tests, or items from physical examination, identified as predictors. The reason for this is not clear.

In contrast to many of the prognostic studies investigating response to treatment for musculoskeletal conditions, the current study used a follow-up period of one year (Beneciuk et al., 2009; Stanton et al., 2010). Post-treatment response was determined at nine weeks, but the longer period was selected for the main outcome assessment as knee OA is a slowly progressing disease. Additionally it could be argued that looking at long-term success is clinically relevant for many conditions, especially given the cost of treatment provision.

Predictors identified in the current study can be compared to those from a study by Weigl et al. (2006) in which patients with hip or knee OA underwent a multimodal inpatient rehabilitation programme. Multivariate analysis using equivalent definitions of responder revealed female sex, marital status (living with partner), following a diet, and low comorbidities to be predictors of successful outcome at six months. Female sex was the only variable identified in both studies. Interestingly, of the eighteen studies retrieved in a recent review, none identified sex as a predictor of treatment response (Stanton et al., 2010). The current findings may be due to stronger influence of sex in knee OA.

Aside from study design differences, the previous study focussed on personal factors, lifestyle risk factors, and psychological status as their independent baseline variables (Weigl et al., 2006), but they did not include any physical examination items. There was only slight overlap between potential predictors investigated in each study, contributing to difference in results. Combining evidence from both studies could inform variable selection in future prognostic investigation of treatment response in knee OA patients.

Clinical implications

Findings from previous prognostic studies reinforce the importance of information gained from history taking in the prediction of response to physical treatment. Psychosocial factors should be addressed and their influence on outcome recognised. There is good evidence that women will respond well to physiotherapy intervention for knee OA. The importance of physical examination may lie in the selection or grading of treatment techniques, in matching treatment to functional limitations, or in monitoring patient progress, rather than as predictors of outcome.

7.7. Strengths and limitations

7.7.1. Strengths

The findings from the current prognostic study are generalisable to other knee OA populations because inclusion/exclusion criteria aimed to limit spectrum bias by recruiting participants at all stages of disease, and from a variety of sources in the community, as well as from hospital clinics. The findings are also generalisable to physiotherapy practice in New Zealand as both assessors and treatment providers were typical of physiotherapists in New Zealand, with a variety of training and years of experience. All assessment techniques and interventions are routinely used in clinical practice. Baseline assessment included readily available patient self-report instruments, as well as simple clinical tests

that did not require any specialist training or equipment. Identified predictors are therefore quick and easy to use in a variety of clinical settings. The absence of any specialised physical tests in the primary model means that medical practitioners, specialist nurses, or physiotherapists should all be able to use it to predict likely response to physiotherapy intervention for patients with knee OA.

The prognostic study was nested within an RCT, both of which were subject to rigorous research procedures including blinding of baseline assessors and outcome assessors. Blinding to group allocation was maintained throughout the analysis. Assessments were standardised and used questionnaires, self-report measures, and clinical tests that had been validated for use with knee OA patients or in older person populations. As there was no sound methodology for assessing and reporting irritability, a new scale was developed with good psychometric properties (Chapter 3). When inter-rater reliability of tests was unclear for this population, it was established in a separate investigation (Chapter 4). Moderate or substantial levels of inter-rater reliability for most variables increased confidence that measures accurately reflected patient condition and not variance between testers. This could assist translation of research findings into clinical practice.

Treatment was standardised as recommended when investigating predictors of treatment response (Beneciuk et al., 2009), although physiotherapists were encouraged to tailor the interventions to individual participants based on impairments and other factors such as irritability or comorbidities (Abbott et al., 2009). Treatment was provided by physiotherapists not involved in outcome assessment. They received training prior to providing interventions and had ready access to a Manual of Operating Procedures (MOP) to ensure consistency (Appendix N).

Use of recognised outcome measures (WOMAC and GRC) should increase confidence in results as well as making them more easily comparable with similar research. Also the use of OMERACT-OARSI response criteria ensured improvement was clinically important and considered information from different domains of pain, function and participant rating of overall improvement (Pham et al., 2004; Dworkin et al., 2008).

7.7.2. Limitations

The inclusion of participants who subsequently underwent joint arthroplasty caused some difficulty with analysis. Joint replacement is not considered as a satisfactory outcome measure in clinical trials (Altman et al., 2005) and could not simply be used to represent “failure” of intervention. Participants in this trial who went forward for surgery included some who had not completed the intervention, others who had benefited from the intervention, and others who were worse despite intervention. Therefore in the primary analysis participants who underwent joint arthroplasty during the course of the trial were counted as “missing data”, which was handled using multiple imputations (MI). This complied with the intention to treat (ITT) principle of the trial (Abbott et al., 2009). However, sensitivity analysis of the modified cohort where joint arthroplasty patients were excluded changed the primary prognostic model (Table 6.29). These factors suggest possible introduction of bias by including participants in the trial who were already candidates for joint surgery (Altman et al., 2005).

The focus on simple clinical tests could have resulted in potential predictors being missed. For example, there is some evidence that immediate response to laterally wedged insoles, baseline disease severity, and baseline function can predict short-term (three month) outcome in patients with knee OA (Hinman et al., 2008). However, assessment in that study included measurements from

radiographs and three-dimensional gait analysis, both of which are beyond the scope of routine clinical practice. Therefore, despite some evidence that immediate response to laterally wedged insoles and baseline disease severity (assessed radiographically) are predictors of outcome for patients with knee OA, neither were included in the current study. Similarly, there was no assessment of gait or related biomechanical features, as assessment of these variables requires sophisticated equipment beyond the reach of routine clinical practice.

A limitation of the current study is the absence of a measure of baseline severity. It has been recommended that disease severity should incorporate measures of pain, function, and structural severity, the latter being assessed radiographically (Gossec et al., 2007). Baseline radiographic severity was identified as a predictor of treatment response in a previous study of patients with knee OA (Hinman et al., 2008). ACR clinical criteria were used for inclusion of participants into the trial and no radiographic investigations were performed as part of the RCT. Initially there was the intention to use previous radiographs to evaluate structural severity (Appendix H p.17). However, not all participants had a recent radiograph and there was marked variation in the methods of reporting. These inconsistencies meant it was not possible to incorporate radiographic structural severity into the analyses. WOMAC was used to assess pain and function, but as it was also the main outcome measure, it was not used as a baseline predictor to avoid generation of artificially high statistical association. Other measures such as pain (NPRS) and symptom duration could be considered as surrogates of baseline severity but, as previously discussed, both are subject to their own limitations.

In addition to not having a measure of disease severity to evaluate as a baseline predictor, it was not possible to state that all participants were recruited at a similar (early) stage of disease. This latter point has been stated as a desirable

feature of prognostic studies and is evaluated as an indicator of quality in systematic reviews of observational studies (Altman and Riley, 2005; Centre for Reviews and Dissemination, 2008). This shortcoming was partly addressed in the analysis as change scores were used to determine outcome.

An additional limitation of this study is the sample size. The treatment cohort of eight-five participants was sufficient for the main prognostic study according to an *a priori* decision for analysis. The largest model contained seven variables, which met the recommended ratio of ten participants to each predictive variable (Childs and Cleland, 2006). However, an alternative approach suggests adequate sample size should be evaluated on the basis of outcome events, using a ratio of ten outcome events for each predictor in the final model (Concato et al., 1993; Moons et al., 2009). There were thirty-five successful outcomes in the treatment cohort, meaning an optimal model should contain three or four predictors at most. Arguably, the primary models in the current study are therefore subject to overfitting, resulting in optimistic predictions and models that are too closely adapted to the data (Royston et al., 2009). Another approach is to calculate sample size based on the number of variables included in the initial baseline assessment as candidate predictors (Beneciuk et al., 2009), an interpretation that also highlights that sample size in the current study was insufficient. Although some authors have suggested that prognostic studies require at least several hundred outcome events (Moons et al., 2009b), this recommendation does have to be balanced with other considerations such as increased costs associated with larger sample sizes, and may be mediated by prevalence of disease or outcome event, severity of consequences of inaccurate prognosis and pre-test probability of outcome (Childs and Cleland, 2006). Many prognostic studies do not discuss sample size or the implications of overfitting models, with a review of CPRs in physiotherapy identifying that only 40% of

studies had an adequate sample size (Beneciuk et al., 2009). This shortcoming should be addressed in future research.

Subgroup analysis was limited due to the sample size. The treatment cohort was subdivided into three treatment groups who received different combinations of treatment (exercise therapy alone, manual therapy alone, or a combination of exercise and manual therapy). Successful outcome was different for each group: exercise therapy was 36%; manual therapy 41%; combined exercise and manual therapy group was 29%. Ideally sub-group analysis would have been performed on the different treatment groups to determine whether different predictors were associated with different interventions. However, the sample size was insufficient for this purpose. It also meant the size of the usual care group was too small to definitively state whether variables were predictors of treatment response or natural progression.

The decision to avoid creating cut-points in data in order to retain as much information as possible was based on recommendations from previous research (Altman et al., 1994; Altman and Royston, 2006; Royston et al., 2006). This decision allowed retention of two continuous baseline variables in the analysis (the i-score, hip external rotator muscle strength) and identification of their role as predictors. However, some of the resulting multivariate models ended up with both dichotomous and continuous predictive variables. This limited further analysis using accuracy statistics and calculation of post-test probabilities for those models, which is an essential step in the development of CPRs.

A related point is that sensitivity analysis introducing a cut-point for age resulted in age (> 57 years) being retained in analysis, as it exhibited a univariate association with successful outcome ($p < 0.08$). There was no association of age with success when it was analysed as a continuous variable, raising the issue that

dichotomising a variable may make it more likely to be selected as a predictor. This opinion requires further research to investigate its validity.

In the current study, outcome was calculated from both the pain and physical function subscales of WOMAC, decreasing reliance on the physical function subscale. The WOMAC has been criticised for the decreased ability of the physical function subscale to detect change due to the overlap of activities with the pain subscale (Stratford and Kennedy, 2004). Despite this it is a widely used outcome measure and has validity and responsiveness in knee and hip OA populations (Bellamy et al., 1988; Angst et al., 2001).

A long duration of follow-up (one year) before outcome assessment in the current study may have led to problems with patient recall for the GRC. The use of GRC as a transitional rating tool relies on participant recall of past health status, which has been identified as a weakness, with longer intervals being at greater risk of bias (Kamper, 2009). This difficulty with recall means participants focus on current status, and while this may inflate ratings of change it is not thought to be a significant source of bias (Guyatt et al., 2002; Schmitt and Di Fabio, 2005). OMERACT-OARSI criteria, with two levels of responder, were used to ensure different domains were included to determine successful outcome, decreasing reliance on the GRC.

The use of an additional post-estimation analysis was a new approach for this type of study and provided valuable insight into the importance of specific variables in the prognostic models. It should be noted this new approach entailed calculation of probability performed immediately after regression. As such it did not incorporate measures of sensitivity, specificity, or LRs, meaning results are not directly comparable with post-test probability of success described following analysis with accuracy statistics and the final prognostic models.

7.8. Summary

7.8.1. Clinical implications

Health practitioners can use baseline variables (posterior knee pain, disturbed sleep, absence of knee injury, longer duration of symptoms, instability, and female sex) to predict likely success with physiotherapy intervention for patients with knee OA. Results are promising that application of a CPR would be able to identify sub-groups of patients with an increased probability of sustained treatment response, although currently this CPR is at an early stage of development. Similarly, the CPR should be able to identify patients unlikely to succeed with physiotherapy and direct them towards other interventions. This will avoid unnecessary or ineffective use of resources, in addition to limiting patient distress and frustration.

Other features of the research relevant to clinical practice are outlined in the previous sections. Identification of predictors will promote their use in clinical practice and highlight their value to assessment of knee OA. It may be necessary to consider the complexity of certain features, such as disturbed sleep, and adjust assessment accordingly. In addition, findings give some reassurance to practitioners that many patient characteristics are not barriers to success with physiotherapy including, increased BMI, older age, malalignment, longer duration of symptoms (> 5 years), disturbed sleep, or knee instability.

The results reinforce the importance of demographic information and details of patient history for making prognostic statements in this patient group. Physical examination findings are less important for prognosis and their value may lie in treatment selection or monitoring progress. Physiotherapists should pay close attention to modifiable predictors and ensure they are addressed by treatment. Additionally, factors such as irritability are important to individual patient response, and emphasise the importance of tailoring treatment to individuals.

It is useful to note that success immediately following completion of a treatment programme (nine week outcome), gives a similar prediction of one year success as use of baseline variables. This is particularly useful in clinical situations where baseline data may not be available or assessment may not have included evaluation of all predictors. There is added value from combining baseline data and nine week response to give greater probability of success at one year. Justification is minimal for providing a course of physiotherapy when baseline predictors suggest it is unlikely to be successful. However, for some patients, alternatives such as joint arthroplasty pose unacceptable risks, are limited by healthcare resource allocation, do not offer substantial prospect of improvement, or are perceived as an unattractive treatment option (Ballantyne et al., 2007). Clinicians must be mindful of patient preference, including desire to avoid surgery or use conservative treatment approaches. In such cases prognostic findings from the current study can be used to give health practitioners and patients realistic information on likely outcome following physiotherapy treatment.

The new method for identifying probability of successful outcome with specific combinations of named predictors should have some clinical appeal. The knowledge that particular combinations of variables offer the likelihood of greater probability of success may further influence health practitioners and patients to opt for physiotherapy treatment. However, this method also increases the complexity of findings that may create a barrier to clinical implementation.

Findings from the current study represent the preliminary stages of CPR development and require further validation and impact analysis (McGinn et al., 2000; Childs and Cleland, 2006). In the absence of any validated CPR for patients with knee OA, this information represents the best evidence currently available.

7.8.2. Research implications

Development of a CPR involves a three step process (McGinn et al., 2000). The first step, undertaken in this study, involves the derivation of the rule intended to identify predictors. This should be followed by a validation step where the rule is tested in different populations. The final step looks at implementation of the rule and the impact it has on clinical practice, including patient outcomes and cost (McGinn et al., 2000; Childs and Cleland, 2006).

There has been much debate in the physiotherapy literature about the methods used to derive a CPR for treatment response (Hancock et al., 2009; Fitzgerald, 2010; Stanton et al., 2010). Although the current study was nested within an RCT, size of the control group (usual care) meant subgroup analysis lacked power to determine definitively if baseline variables were predicting treatment response or natural progression over one year. Despite limited evidence that they were predicting treatment response, findings are equivalent to those obtained in a single-arm trial. However the study has served to reduce the number of possible predictors, which could be examined in a future, appropriately powered, RCT. If conducted in different knee OA populations, the subsequent study could also provide preliminary validation of the CPR (Fitzgerald, 2010).

Further development of the CPR would require replication of the current study with increased sample size. The difficulty in estimating sample size for prognostic studies has been recognised (Moons et al., 2009b), with one option being to base the sample size on the number of candidate predictor variables, rather than the number in the final model. However, recent recommendations reiterate previous suggestions that sample size should be related to outcomes, at a ratio of ten outcome events to each predictor in the final model (Concato et al.,

1993; Moons et al., 2009). Additionally sample size should be sufficient to allow sub-group analysis of different treatments.

Predictors from all models in the current study should be considered for inclusion in future research, as well as those identified in a similar study (Weigl et al., 2006). Inclusion of other baseline variables with recognised clinical importance to knee OA (e.g. muscle strength, pain, malalignment and BMI), could justifiably be included in future study, especially variables such as malalignment which require development of a valid method of clinical assessment. Additionally, possible predictors beyond the scope of the current study could be included, especially if they have potential to be modified by the described intervention and have published evidence of an association with outcome.

Biomechanical factors are a promising source of predictors as there is some evidence of short-term treatment response in patients with knee OA (Hinman et al., 2008). Detailed biomechanical analysis of gait, assessment of alignment, knee adduction force, and hip abduction force require sophisticated equipment and analysis that is uncommon in routine clinical practice. However, their potential as predictors could be examined in a research environment. Any positive findings might then support future research to develop valid and reliable measures of biomechanical features with clinical utility.

Adding patient perspective is a recommended approach to outcome assessment. The WOMAC index has been individualised by getting patients to nominate five items from the physical function subscale they consider important for measuring improvement in hip or knee OA (Seror et al., 2008). This method has good psychometric properties and responsiveness, better reflecting patient concerns, and could be used in future studies as a modified method of evaluating outcome with WOMAC (Seror et al., 2008).

An alternative to GRC, as an externally anchored method for evaluating clinically relevant change, is the patient acceptable symptom state (PASS) (Tubach et al., 2005). As a tool it deals with patient satisfaction with their current state, avoiding the issues of recall and other problems associated with GRC. The PASS may be of particular value in knee OA research where outcomes are measured over longer periods.

Development of this CPR or other similar studies will need to address some of the shortcomings of the current statistical analysis. These centre on the decision not to create cut-points in data, leading to a mix of continuous and categorical or dichotomised variables. Although this had the advantage of retaining data, it led to complications in the analysis, hindering the use of accuracy statistics and the development of a CPR when continuous variables were identified as predictors (the i-score, hip external rotator muscle strength). If the aim of future research is to develop a CPR, rather than solely to identify predictors of treatment response, variables must be dichotomised prior to regression analysis using a non-data driven approach such as receiver operating characteristic (ROC) curve analysis.

Post-estimation analysis resulting in probabilities for specific combinations of predictors requires validation in future studies. It may be a useful adjunct to current methods of analysis by highlighting the importance of specific predictors, allowing development of targeted therapy. It may also raise questions relating to subgroups of treatment responders, such as why women appear to respond better to treatment, that can be addressed in future research. However, the level of detail provided needs to be balanced with increasing complexity of prognostic information, which could be too detailed or time consuming for routine clinical use.

The appeal of prognostic models to clinicians should be investigated as part of translating research findings into practice. Impact analysis is the third step in

CPR development and follows on from validation of the rule (McGinn et al., 2000). However, some aspects relating to implementation of prognostic findings could be investigated in advance and could potentially improve clinical uptake. Issues such as the amount of change in probability considered important or relevant could be investigated, accepting that this level may change depending on the perspective of the stakeholder i.e. health funder, health practitioner, or patient. Also, a future study could examine the degree of awareness and understanding amongst physiotherapists about CPRs. Their knowledge and attitudes about the interpretation and implementation of prognostic findings could be investigated, in addition to identifying potential barriers to uptake. For example, simpler models that are easier to remember may be more clinically appealing. This information may influence the extent to which a CPR is developed, as well as its complexity.

8. Predictors of outcome following physiotherapy for individuals with knee osteoarthritis: Conclusions

The increasing burden of knee osteoarthritis (OA) and the growing body of research showing physiotherapy as an effective method of managing that burden, provide an important context for the studies included in this thesis.

Heterogeneity in clinical presentation of the disease, and recognition that not all patients with knee OA respond in a similar manner to physiotherapy treatment, led to the current investigation.

This chapter will provide an overview of the thesis with a summary of the main findings of the prognostic study. The main findings of the supplementary investigations and their contribution to the prognostic study will be highlighted. The impact of the findings on existing knowledge will be outlined, including the contribution to prognostic research, specifically in the area of treatment response to physiotherapy. Strengths and limitations, research and clinical implications have been discussed in previous chapters and sections of the thesis. Issues related to the whole body of work will be considered in this chapter.

8.1. Summary of major findings of prognostic study

The main aim of the thesis was to identify predictors of good outcome at one year following physiotherapy treatment for patients with knee OA. This represented the derivation phase of a clinical prediction rule (CPR) to identify subgroups of treatment responders with increased likelihood of success with physiotherapy.

The *a priori* hypotheses can be accepted with the identification of six baseline predictors of good outcome: posterior knee pain, disturbed sleep, absence of previous knee injury, instability, longer duration of symptoms, and female sex.

Secondly post-test probability was altered by the presence or absence of clusters of predictors. A combination of at least four out of six predictors increased likely success from 35% to 66%, while presence of two or less predictors decreased likely success from 35% to 11%. All variables were assessed at baseline examination using tests and measures commonly used in routine clinical practice, and provide health practitioners and patients with useful information for making clinical decisions.

A new approach for CPRs was developed following multivariate regression, which involved using a post-estimation test. Probability was calculated for specific combinations of named variables, enriching the information usually provided in CPRs. Hence post-test probability of 66% for successful response, with at least four out of six predictors, could range from 56% to 87% depending on the actual combination of four or five variables. Disturbed sleep, absence of knee injury, symptom duration (> 5 years), and female sex gave a probability of 56%, while posterior knee pain, disturbed sleep, absence of previous knee injury, instability, and female sex, produced probability of 87%, yet both conditions met the CPR of at least four out of six predictors present. This approach could be adopted for other CPRs to maximise increases in post-test probability and help identify the most important predictors of outcome.

The new approach helped confirm immediate successful response to the treatment programme (nine week response), added to baseline variables, and produced another CPR with greater change in post-test probability. In isolation, nine week response increased probability of success at one year to 52%. This finding is important because full information from baseline assessment is not always available. However, when baseline predictors and nine week response were combined, the optimal model of at least five out of seven predictors increased likely success to 86%. The new post-estimation analysis confirmed nine

week success was an important contributor to this result, with probability of success ranging from 73% to 93% when it was included in the model.

These findings add significantly to the field of knee OA conservative management, by identifying six predictors of successful response to physiotherapy intervention. Of these, only female sex had previously been identified as a predictor of response following inpatient rehabilitation (Weigl et al., 2006). Following sensitivity analysis, additional baseline variables with limited evidence as predictors of success were the i-score and hip external rotator muscle strength. Findings from the current and previous studies highlight a range of variables for consideration in future prognostic research.

Baseline variables identified as predictors of treatment success are also important for other aspects of knee OA patient assessment and management: for example, to guide technique selection or as targets of intervention. These are factors which should be considered in the design of new or different therapeutic approaches.

This study has added further information to prognostic research in musculoskeletal conditions, an area of increasing importance in physiotherapy (Beneciuk et al., 2009; Fritz, 2009). Recommended methods of study design and analysis were used (Laupacis et al., 1997; Royston et al., 2009), creating a sound platform from which to further develop this CPR. Findings should be regarded with some caution as they represent a preliminary stage of CPR development and await validation and impact analysis before they can be used with confidence in clinical practice (McGinn et al., 2000; Childs and Cleland, 2006; Moons et al., 2009a). Specifically the study needs to be replicated with a larger sample size as the models with more than three or four variables are arguably over fitted, increasing the likelihood of spurious findings. Additionally sensitivity analysis revealed findings were subject to change following alterations in some

parameters and methodology, indicating further research is required to develop a CPR that is more robust.

8.2. Summary of major findings of supplementary studies

Three supplementary studies informed conduct of the main prognostic investigation: the systematic review of the literature (Chapter 2); the irritability study (Chapter 3); and the investigation into reliability of tests and measures used in the baseline assessment (Chapter 4).

8.2.1. Systematic review identifying predictors of progression in patients with knee OA

Variables identified with strong evidence as predictors of knee OA progression were: varus alignment, presence of OA in multiple joints, age, radiographic features, and body mass index (BMI) (over a longer period > 3 years). Additionally participation in physical activity was not associated with progression. Raising the threshold of study quality revealed strong evidence that presence of baseline pain or functional limitation were not associated with progression. There were numerous other potential predictors where evidence was limited or conflicting.

Variables included in the baseline assessment of the main prognostic study were those with strong evidence from a number of studies. Variables with reported large effect sizes were considered for inclusion, even if overall evidence for prediction of progression was limited or conflicting. Also included were potentially modifiable variables where evidence on progression was limited or conflicting.

The focus of the systematic review was delimited to predictors of progression of knee OA for which the methods of assessment were easily utilised in clinical practice. The findings were therefore well-aligned with the main prognostic study.

8.2.2. The concept of irritability and development of the i-score

Evidence from previous literature and a qualitative (Nominal Group Technique) approach supported the view that irritability is an important component of symptom behaviour that influences physiotherapists' selection and application of assessment and treatment techniques, and may be associated with response to treatment (Maitland, 1986, 1991; Hurley et al., 2002; Smart and Doody, 2007).

A review of the literature indicated that there was no standardised method of reporting and recording irritability for either clinical or research purposes. The i-score was therefore developed to provide a tool for the current research. The i-score has four items: aggravating activity, time to onset or worsening of pain, intensity of pain, and time for return to resting level of pain. These were categorised and ordered to reflect increasing severity. Adding the scores gave a maximum score of forty, with a higher score representing higher irritability.

Development of the i-score achieved its objectives by providing a standardised method of reporting with "moderate" psychometric properties. Thus, irritability could be included as a potential predictor of outcome in the main prognostic study.

Although the i-score had a univariate association with good outcome at one year, it was not retained in any multivariate models; with the exception of a stratified sensitivity analysis (total joint arthroplasty patients excluded), when male sex and the i-score were predictors of success at one year. This finding lacks clinical utility as the model involved several stages of analysis, making interpretation more complex. However it provides justification for including the i-score as a potential predictor in futures studies of treatment response in patients with knee OA.

The i-score can be used with confidence in clinical practice to assess irritability in patients with hip or knee OA. It will enhance communication between physiotherapists, and facilitate teaching, thus improving patient care and outcomes. There is also potential for it to be further developed for use in other musculoskeletal conditions.

8.2.3. Inter-rater reliability of clinical tests used to evaluate patients with knee OA

Results from both the review of literature and the investigation performed, showed that the majority of tests had at least moderate levels of inter-rater reliability. Tests based on patient history or symptom report had some of the strongest results. Range of motion testing and physical performance measures (PPM) also performed well. Some of the weakest reliability results were for palpatory tests or measures of alignment.

Inter-rater reliability is only one factor to be considered for screening clinical measures for inclusion in the baseline assessment. Importance of the characteristic being measured to the management of knee OA must also be considered. For example, observation of valgus/varus alignment of the knee showed “fair” level of agreement ($\kappa = 0.3$) (Landis and Koch, 1977) but was still included in the baseline assessment for the prognostic study. This was due to the strong evidence for varus malalignment as a predictor of progression of knee OA (Chapter 2) and recognition of a complex interplay with other modifiable factors such as BMI and quadriceps strength (Sharma et al., 2003b; Lim et al., 2008; Niu et al., 2009).

A reasonable level of inter-rater reliability allows clinicians to be confident that their test results match those of the study and increases confidence in treatment decisions based on those tests (Hicks et al., 2003). The inter-reliability results provide new information about reliability of clinical tests for use with elderly patients, or in hip and knee OA populations.

8.3. Strengths and limitations

Study procedures were typical of routine physiotherapy practice in New Zealand, with New Zealand registered physiotherapists providing the interventions. Consequently the findings are generalisable to physiotherapists practising in New Zealand or similar health care environments.

Additionally, findings from the reliability study, the irritability study, and the main prognostic study are all generalisable to other populations of knee OA patients due to the wide spectrum of disease included in the study sample. This was partly due to the use of American College of Rheumatology (ACR) clinical criteria to classify patients as having knee OA (Altman et al., 1986). The intent was to include patients in the early stages of knee OA who had not yet developed radiographic change, as well as those with established or severe disease. It has been suggested that physiotherapy or other non-pharmacological interventions may be more effective at early stages of knee OA disease (Fitzgerald and Oatis, 2004; Fransen, 2004).

A difficulty with using clinical criteria is posed by some evidence that diagnosis of early knee OA on the basis of clinical features was not necessarily associated with development of incident radiographic disease at three years (Peat et al., 2010); however, evidence from another study showed that 86% of participants with knee pain at baseline had developed tibio-femoral OA by twelve years (Thorstensson et al., 2009a). It is therefore possible that some misclassification of participants with knee OA in the current study could have led to inclusion of individuals without the disease.

Another problem with using ACR criteria to classify patients as having knee OA is the lack of a clear operational definition for presence of knee pain in the original paper (Altman et al., 1986). This means it can be interpreted in a variety of ways; i.e. any history of knee pain, or knee pain in the last week, or knee pain

in the last month, or knee pain on most days and so forth. This inconsistency will influence patient selection and sample characteristics, and could potentially contribute to discrepancies in findings between different studies.

In this study, there was no upper limit on disease severity and participants were not excluded if they had advanced disease or if they were on a waiting list for joint arthroplasty surgery (Abbott et al., 2009). This was problematic as 15% of the knee OA participants had joint arthroplasty within the one year follow-up period. Decisions for handling the data for these participants have been explained in previous chapters, with the sensitivity analysis revealing that primary prognostic models were changed when joint arthroplasty participants were excluded from the analysis.

8.4. Research implications

As previously mentioned this CPR requires validation in a different population of knee OA patients, and preferably with different clinicians providing the interventions. Sample size should be increased to address possible overfitting of the model and to allow for complete sub-group analysis, as it is possible there are different predictors for different types of treatment or for no treatment. The use of clinical criteria, either ACR clinical criteria or EULAR recommended clinical criteria (Altman et al., 1986; Zhang et al., 2010) will ensure individuals at an early stage of knee OA disease are included in future studies. Additionally, exclusion criteria could be adjusted to limit recruitment of patients likely to undergo joint arthroplasty during the study period.

Presence of knee pain is an important component of clinical criteria for classification or diagnosis of knee OA and warrants further research to clarify the optimal method of assessment. A review of the current variation in operational definitions of “presence of knee pain” could be followed by recommendations for a standardised question to be utilised in future studies.

Patient perspective and understanding of prognosis should also be investigated, as a goal of this research is better information for both health practitioners and patients to assist decision-making for clinical management of knee OA. In addition to identifying the best methods of education and dissemination of information, research could examine the amount of change in probability of success that is meaningful for patients.

There is evidence that older patients with musculoskeletal pain are interested in prognostic information, but it is not commonly included in consultations with General Practitioners (GPs) (Mallen and Peat, 2009). Time limitation was the reason many GPs did not fully discuss prognosis with this patient group (Mallen et al., 2007a). However, one reason patients gave for wanting prognostic information was to alter activity to prevent deterioration, while reasons for *not* wanting information included the perception that nothing could be done, and that progression was inevitable (Mallen and Peat, 2009). The clinical encounter therefore gives plenty of opportunity to challenge misconceptions about knee OA and to direct patients to appropriate intervention. However, recognising the limited time available to achieve this during a GP consultation should encourage investigation of other ways to deliver the information. Physiotherapists may be well placed to do this, as their clinical encounters with patients tend to be of longer duration and repeated over a period of weeks. This provides a better opportunity for them to deliver educational and prognostic information. The extent to which physiotherapists currently include prognosis in their usual clinical practice is unknown.

A related area of research stems from the evidence that around 50% of patients with severe and disabling knee pain do not consult their GPs (Jordan et al., 2006; Mitchell et al., 2006; Bedson et al., 2007). Furthermore, adherence to treatment guidelines by health practitioners is inconsistent, and clinical

management of knee OA does not always comply with the best evidence (DeHaan et al., 2007; Jamtvedt et al., 2008b; Poitras et al., 2010). The combined effect is that non-pharmacological interventions are not implemented, or are done so on a self-care basis by the patient (Porcheret et al., 2007); thus opportunities are lost for effective clinical management to influence disease progression. The optimal methods of altering patient and health care practitioner perceptions about inevitability of knee OA progression and the likely benefits of physiotherapy treatment should be the subject of future research. Additionally, while there is some evidence for the reasons patients do or do not consult their GP about knee OA, there is little evidence about their reasons for consulting a physiotherapist, or the factors that would influence a GP to refer patients for physiotherapy intervention.

8.5. Clinical implications

It has been emphasised that CPRs, such as the one developed here, should not be applied in clinical practice until fully validated (McGinn et al., 2000; Childs and Cleland, 2006; Moons et al., 2009a). However, findings from this thesis could be used to persuade clinicians to re-examine their practice. With the knowledge that a proportion of patients with knee OA are unlikely to consult a health practitioner, they should be alert to signs and symptoms, perhaps when patients present for other health care problems, and be proactive about suggesting physiotherapy intervention, especially if predictors identified in this study are identified during assessment. Information can be used to educate patients about the likely benefit of treatment, stating that 35-40% of individuals with knee OA are likely to be significantly better at one year following this particular physiotherapy intervention. Additionally this proportion may be increased in the presence of identified baseline characteristics, although caution should be exercised about claiming specific probabilities of success.

Design of the prognostic study involved a detailed examination of the components included in a physiotherapy assessment, which is an extensive and often time-consuming process. Inclusion of items in any assessment should have sound clinical justification. Findings from the thesis have identified variables predictive of treatment response and of natural progression over the course of one year. Other variables (or items) in the assessment may be diagnostic; others direct assessment and treatment choices; others record patient distress or level of discomfort; while others may be used to assess outcome or indicate the need for progression of treatment. Assessment in physiotherapy is different to medical practice in that it frequently takes place over several sessions, allowing information gathered, including response to techniques, to influence diagnosis and patient management. These advantages should be exploited by physiotherapists to refine clinical decisions and maximise opportunities to educate patients and influence lifestyle choices, issues which have been identified in this thesis as important for managing patients with knee OA.

Beyond the specifics of this CPR for predicting treatment outcome for patients with knee OA, physiotherapists should consider the importance of prognostic information in general to their practice, and to their patients.

8.6. Conclusion

The recognition that physiotherapy is effective for managing some but not all patients with knee OA led to the primary aim of this thesis: to identify predictors of success at one year for patients with knee OA following physiotherapy treatment. The four studies comprising this thesis contributed to the major finding that presence of at least four out of six predictors increased probability of successful outcome from 35% to 66%. The presence of two or less predictors decreased probability of success to 11%. Adding immediate post-treatment programme success to the model increased the probability of one year

success even further to 86% when at least five out of seven predictors were present.

The previous lack of research into the area of predictors of treatment response in patients with knee OA reinforces the importance of these findings. The development of this programme of prognostic research utilised existing knowledge and guidelines, but also critiqued methods and examined ways of improving or developing them. Clinical relevance was the main driver to add more detail to the CPR, identifying specific combinations of named predictors which changed estimates of probability for the different models.

Understanding limitations of the study has indicated a clear route forwards for ongoing study, specifically validation in a different sample with a larger sample size, and impact analysis of the CPR prior to clinical implementation. The prognostic models were not robust; they were sensitive to changes in analysis, suggesting further development is required. Other areas for future research include exploring physiotherapists', patients' and other health practitioners' knowledge of and attitudes to prognostic information, and identification of possible barriers to implementation.

Findings from the supplementary studies, in addition to assisting the prognostic research, provided some useful independent information by identifying current strength of evidence for predictors of natural progression of knee OA; developing a tool to measure irritability; and increasing knowledge on the inter-rater reliability of clinical tests. These areas offer some opportunities for future research, especially in the area of scale development.

Clinical implications are important and have been highlighted throughout the thesis. Changing clinical behaviour on the basis of research findings is a challenge, but having this endpoint in mind influenced many decisions such as inclusion of simple and easily accessible tests and measures, provision of an

easily reproducible treatment programme, and inclusion of a broad spectrum of knee OA disease in the study participants. Clinical practice has informed research decisions and the goal is that research findings will now inform clinical practice.

References

- ABAD, V. C., SARINAS, P. S. A. & GUILLEMINAULT, C. 2008. Sleep and rheumatologic disorders. *Sleep Medicine Reviews*, 12, 211-228.
- ABBATE, L. M., STEVENS, J., SCHWARTZ, T. A., RENNER, J. B., HELMICK, C. G. & JORDAN, J. M. 2006. Anthropometric measures, body composition, body fat distribution, and knee osteoarthritis in women. *Obesity*, 14, 1274-81.
- ABBOTT, J., CLARE, M., MCKENZIE, J., DAVID, G., THEIS, J. & CAMPBELL, A. 2009. Exercise therapy, manual therapy, or both, for osteoarthritis of the hip or knee: A factorial randomised controlled trial protocol. *Trials*, 10.
- ACCESS ECONOMICS 2005. The economic cost of arthritis in New Zealand. Canberra: Arthritis New Zealand.
- ACCESS ECONOMICS 2010. The economic cost of arthritis in New Zealand in 2010. Canberra: Arthritis New Zealand.
- ALTMAN, D. G. 2001. Systematic reviews of evaluations of prognostic variables. *BMJ*, 323, 224-8.
- ALTMAN, D. G., LAUSEN, B., SAUERBREI, W. & SCHUMACHER, M. 1994. Dangers of using "optimal" cutpoints in the evaluation of prognostic factors. *Journal of the National Cancer Institute*, 86, 829-35.
- ALTMAN, D. G. & RILEY, R. D. 2005. Primer: an evidence-based approach to prognostic markers. *Nature Clinical Practice Oncology*, 2, 466-472.
- ALTMAN, D. G. & ROYSTON, P. 2006. The cost of dichotomising continuous variables. *BMJ*, 332, 1080.
- ALTMAN, R., ASCH, E., BLOCH, D., BOLE, G., BORENSTEIN, D., BRANDT, K., CHRISTY, W., COOKE, T. D., GREENWALD, R., HOCHBERG, M. & ET AL. 1986. Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Association. *Arthritis & Rheumatism*, 29, 1039-49.
- ALTMAN, R. D., ABADIE, E., AVOUAC, B., BOUVENOT, G., BRANCO, J., BRUYERE, O., CALVO, G., DEVOGELAER, J. P., DREISER, R. L., HERRERO-BEAUMONT, G., KAHAN, A., KREUTZ, G., LASLOP, A., LEMMEL, E. M., MENKES, C. J., PAVELKA, K., VAN DE PUTTE, L., VANHAELST, L., REGINSTER, J. Y., GROUP FOR RESPECT OF, E. & ETHICS IN, S. 2005. Total joint replacement of hip or knee as an outcome measure for structure modifying trials in osteoarthritis. *Osteoarthritis & Cartilage*, 13, 13-9.
- AMIN, S., BAKER, K., NIU, J., CLANCY, M., GOGGINS, J., GUERMAZI, A., GRIGORYAN, M., HUNTER, D. & FELSON, D. 2009. Quadriceps strength

- and the risk of cartilage loss and symptom progression in knee osteoarthritis. *Arthritis and Rheumatism*, 60, 189-198.
- AMIN, S., GUERMAZI, A., LAVALLEY, M. P., NIU, J., CLANCY, M., HUNTER, D. J., GRIGORYAN, M. & FELSON, D. T. 2008. Complete anterior cruciate ligament tear and the risk for cartilage loss and progression of symptoms in men and women with knee osteoarthritis. *Osteoarthritis and Cartilage*, 16, 897-902.
- AMIN, S., NIU, J., GUERMAZI, A., GRIGORYAN, M., HUNTER, D. J., CLANCY, M., LAVALLEY, M. P., GENANT, H. K., FELSON, D. T., AMIN, S., NIU, J., GUERMAZI, A., GRIGORYAN, M., HUNTER, D. J., CLANCY, M., LAVALLEY, M. P., GENANT, H. K. & FELSON, D. T. 2007. Cigarette smoking and the risk for cartilage loss and knee pain in men with knee osteoarthritis. *Annals of the Rheumatic Diseases*, 66, 18-22.
- ANGST, F., AESCHLIMANN, A., MICHEL, B. A. & STUCKI, G. 2002. Minimal clinically important rehabilitation effects in patients with osteoarthritis of the lower extremities. *Journal of Rheumatology*, 29, 131-8.
- ANGST, F., AESCHLIMANN, A., STEINER, W. & STUCKI, G. 2001. Responsiveness of the WOMAC osteoarthritis index as compared with the SF-36 in patients with osteoarthritis of the legs undergoing a comprehensive rehabilitation intervention. *Annals of the Rheumatic Diseases*, 60, 834-40.
- AXFORD, J., HERON, C., ROSS, F. & VICTOR, C. R. 2008. Management of knee osteoarthritis in primary care: pain and depression are the major obstacles. *Journal of Psychosomatic Research*, 64, 461-7.
- BALLANTYNE, P. J., GIGNAC, M. A. M. & HAWKER, G. A. 2007. A patient-centered perspective on surgery avoidance for hip or knee arthritis: lessons for the future. *Arthritis & Rheumatism*, 57, 27-34.
- BANDY, W. D. & IRION, J. M. 1994. The Effect of Time on Static Stretch on the Flexibility of the Hamstring Muscles. *Physical Therapy*, 74, 845-852.
- BARAKATT, E. T., ROMANO, P. S., RIDDLE, D. L. & BECKETT, L. A. 2009a. The reliability of Maitland's irritability judgements in patients with low back pain. *The Journal of Manual & Manipulative Therapy*, 17, 135-140.
- BARAKATT, E. T., ROMANO, P. S., RIDDLE, D. L., BECKETT, L. A. & KRAVITZ, R. 2009b. An exploration of Maitland's concept of pain irritability in patients with low back pain. *The Journal of Manual & Manipulative Therapy*, 17, 196-205.
- BEATTIE, P. F. & NELSON, R. M. 2006. Clinical prediction rules: what are they and what do they tell us? *Australian Journal of Physiotherapy*, 52, 157-63.
- BEATTIE, P. F. & NELSON, R. M. 2007. Evaluating research studies that address prognosis for patients receiving physical therapy care: a clinical update. *Physical Therapy*, 87, 1527-35.

- BEATTIE, P. F. & NELSON, R. M. 2008. Preserving the quality of the patient-therapist relationship: an important consideration for value-centered physical therapy care. *Journal of Orthopaedic & Sports Physical Therapy*, 38, 34-5.
- BEDSON, J. & CROFT, P. R. 2008. The discordance between clinical and radiographic knee osteoarthritis: a systematic search and summary of the literature. *Bmc Musculoskeletal Disorders*, 9, 116.
- BEDSON, J., MOTTRAM, S., THOMAS, E. & PEAT, G. 2007. Knee pain and osteoarthritis in the general population: what influences patients to consult? *Family Practice*, 24, 443-53.
- BELLAMY, N., BUCHANAN, W. W., GOLDSMITH, C. H., CAMPBELL, J. & STITT, L. W. 1988. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *Journal of Rheumatology*, 15, 1833-40.
- BELO, J. N., BERGER, M. Y., REIJMAN, M., KOES, B. W. & BIERMA-ZEINSTRAS, S. M. A. 2007. Prognostic factors of progression of osteoarthritis of the knee: a systematic review of observational studies. *Arthritis & Rheumatism*, 57, 13-26.
- BENECIUK, J. M., BISHOP, M. D. & GEORGE, S. Z. 2009. Clinical Prediction Rules for Physical Therapy Interventions: A Systematic Review. *Physical Therapy*, 89, 114-124.
- BENICHOU, O. D., HUNTER, D. J., NELSON, D. R., GUERMAZI, A., ECKSTEIN, F., KWON, K., MYERS, S. L., WIRTH, W. & DURYEA, J. 2010. One-year change in radiographic joint space width in patients with unilateral joint space narrowing: data from the Osteoarthritis Initiative. *Arthritis Care & Research (2151464X)*, 62, 924-931.
- BENNELL, K. & HINMAN, R. 2005. Exercise as a treatment for osteoarthritis. *Current Opinion in Rheumatology*, 17, 634-40.
- BENNELL, K., HUNT, M., WRIGLEY, T., HUNTER, D., MCMANUS, F., HODGES, P., LI, L. & HINMAN, R. 2010. Hip strengthening reduces symptoms but not knee load in people with medial knee osteoarthritis and varus malalignment: A randomised controlled trial. *Osteoarthritis and Cartilage*, 18, 621-628.
- BENNELL, K. L., HINMAN, R. S., CROSSLEY, K. M., METCALF, B. R., BUCHBINDER, R., GREEN, S. & MCCOLL, G. 2004. Is the Human Activity Profile a useful measure in people with knee osteoarthritis? *Journal of Rehabilitation Research & Development*, 41, 621-30.
- BENNELL, K. L., HUNT, M. A., WRIGLEY, T. V., HUNTER, D. J. & HINMAN, R. S. 2007. The effects of hip muscle strengthening on knee load, pain, and

- function in people with knee osteoarthritis: a protocol for a randomised, single-blind controlled trial. *Bmc Musculoskeletal Disorders*, 8, 121.
- BERTHIAUME, M. J., RAYNAULD, J. P., MARTEL-PELLETIER, J., LABONTE, F., BEAUDOIN, G., BLOCH, D. A., CHOQUETTE, D., HARAOU, B., ALTMAN, R. D., HOCHBERG, M., MEYER, J. M., CLINE, G. A. & PELLETIER, J. P. 2005. Meniscal tear and extrusion are strongly associated with progression of symptomatic knee osteoarthritis as assessed by quantitative magnetic resonance imaging. *Annals of the Rheumatic Diseases*, 64, 556-63.
- BIJLSMA, J. W. J., BERENBAUM, F. & LAFEBER, F. P. J. G. 2011. Osteoarthritis: an update with relevance for clinical practice. *Lancet*, 377, 2115-26.
- BLACK, N. 2006. Consensus development methods. In: POPE, C. & MAYS, N. (eds.) *Qualitative research in health care*. 3rd edition ed. Oxford: Blackwell publishing ltd.
- BLAGOJEVIC, M., JINKS, C., JEFFERY, A. & JORDAN, K. P. 2010. Risk factors for onset of osteoarthritis of the knee in older adults: a systematic review and meta-analysis. *Osteoarthritis & Cartilage*, 18, 24-33.
- BLAND, J. M. & ALTMAN, D. G. 1986. Statistical methods for assessing agreement between two methods of clinical measurement. [Reprint in *Int J Nurs Stud*. 2010 Aug;47(8):931-6; PMID: 20430389]. *Lancet*, 1, 307-10.
- BLAND, J. M. & ALTMAN, D. G. 1999. Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, 8, 135-60.
- BLAND, J. M. & ALTMAN, D. G. 2002. Validating scales and indexes. *BMJ*, 324, 606-607.
- BOHANNON, R. W. 2001. Measuring knee extensor muscle strength. *American Journal of Physical Medicine & Rehabilitation*, 80, 13-8.
- BOSOMWORTH, N. J. 2009. Exercise and knee osteoarthritis: benefit or hazard? *Canadian Family Physician*, 55, 871-8.
- BROSSEAU, L., BALMER, S., TOUSIGNANT, M., O'SULLIVAN, J. P., GOUDREAULT, C., GOUDREAULT, M. & GRINGRAS, S. 2001. Intra- and intertester reliability and criterion validity of the parallelogram and universal goniometers for measuring maximum active knee flexion and extension of patients with knee restrictions. *Archives of Physical Medicine & Rehabilitation*, 82, 396-402.
- BROUWER, G. M., VAN TOL, A. W., BERGINK, A. P., BELO, J. N., BERNSEN, R. M. D., REIJMAN, M., POLS, H. A. P. & BIERMA-ZEINSTR, S. M. A. 2007. Association between valgus and varus alignment and the development and progression of radiographic osteoarthritis of the knee. *Arthritis & Rheumatism*, 56, 1204-11.
- BRUYERE, O., HONORE, A., ETHGEN, O., ROVATI, L. C., GIACOVELLI, G., HENROTIN, Y. E., SEIDEL, L. & REGINSTER, J. Y. L. 2003. Correlation

- between radiographic severity of knee osteoarthritis and future disease progression. Results from a 3-year prospective, placebo-controlled study evaluating the effect of glucosamine sulfate. *Osteoarthritis and Cartilage*, 11, 1-5.
- CENTRE FOR REVIEWS AND DISSEMINATION, U. O. Y. 2008. *Systematic reviews: CRD's guidance for undertaking reviews in healthcare*, York, CRD, University of York.
- CHANG, A., HAYES, K., DUNLOP, D., SONG, J., HURWITZ, D., CAHUE, S. & SHARMA, L. 2005. Hip abduction moment and protection against medial tibiofemoral osteoarthritis progression. *Arthritis & Rheumatism*, 52, 3515-3519.
- CHILDS, J. D. & CLELAND, J. A. 2006. Development and application of clinical prediction rules to improve decision making in physical therapist practice. *Physical Therapy*, 86, 122-31.
- CHO, H., CHANG, C., YOO, J., KIM, S. & KIM, T. 2010. Gender differences in the correlation between symptom and radiographic severity in patients with knee osteoarthritis. *Clinical Orthopaedics and Related Research*, 468, 1749-1758.
- CIBERE, J., BELLAMY, N., THORNE, A., ESDAILE, J. M., MCGORM, K. J., CHALMERS, A., HUANG, S., PELOSO, P., SHOJANIA, K., SINGER, J., WONG, H. & KOPEC, J. 2004. Reliability of the knee examination in osteoarthritis: effect of standardization. *Arthritis & Rheumatism*, 50, 458-68.
- CIBERE, J., THORNE, A., BELLAMY, N., GREIDANUS, N., CHALMERS, A., MAHOMED, N., SHOJANIA, K., KOPEC, J. & ESDAILE, J. M. 2008. Reliability of the hip examination in osteoarthritis: effect of standardization. *Arthritis & Rheumatism*, 59, 373-81.
- CICUTTINI, F. M., JONES, G., FORBES, A. & WLUKA, A. E. 2004. Rate of cartilage loss at two years predicts subsequent total knee arthroplasty: a prospective study. *Annals of the Rheumatic Diseases*, 63, 1124-1127.
- CLAPIS, P. A., DAVIS, S. M. & DAVIS, R. O. 2008. Reliability of inclinometer and goniometric measurements of hip extension flexibility using the modified Thomas test. *Physiotherapy Theory And Practice*, 24, 135-141.
- CLELAND, J. A. 2005. *Orthopaedic clinical examination: an evidence-based approach for physical therapists*, Carlstadt, New Jersey, Icon learning systems.
- CLELAND, J. A., FRITZ, J. M., WHITMAN, J. M. & HEATH, R. 2007. Predictors of short-term outcome in people with a clinical diagnosis of cervical radiculopathy. *Physical Therapy*, 87, 1619-32.
- CONCATO, J., FEINSTEIN, A. R. & HOLFORD, T. R. 1993. The risk of determining risk with multivariable models. *Annals of Internal Medicine*, 118, 201-10.
- COOPER, C., SNOW, S., MCALINDON, T. E., KELLINGRAY, S., STUART, B., COGGON, D., DIEPPE, P. A., COOPER, C., SNOW, S., MCALINDON, T.

- E., KELLINGRAY, S., STUART, B., COGGON, D. & DIEPPE, P. A. 2000. Risk factors for the incidence and progression of radiographic knee osteoarthritis. *Arthritis & Rheumatism*, 43, 995-1000.
- COTTRELL, E., RODDY, E. & FOSTER, N. 2010. The attitudes, beliefs and behaviours of GPs regarding exercise for chronic knee pain: a systematic review. *BMC family practice*, 11.
- CURRIER, L. L., FROEHLICH, P. J., CAROW, S. D., MCANDREW, R. K., CLIBORNE, A. V., BOYLES, R. E., MANSFIELD, L. T. & WAINNER, R. S. 2007. Development of a clinical prediction rule to identify patients with knee pain and clinical evidence of knee osteoarthritis who demonstrate a favorable short-term response to hip mobilization. *Physical Therapy*, 87, 1106-19.
- CUSHNAGHAN, J., COOPER, C., DIEPPE, P., KIRWAN, J., MCALINDON, T. & MCCRAE, F. 1990. Clinical assessment of osteoarthritis of the knee. *Annals of the Rheumatic Diseases*, 49, 768-70.
- CYRIAX, J. 1982. *Textbook of orthopaedic medicine*, London, Balliere Tindall.
- DAHLE, L. K., MUELLER, M., DELITTO, A. & DIAMOND, J. E. 1991. Visual Assessment of Foot Type and Relationship of Foot Type to Lower Extremity Injury. *Journal of Orthopaedic & Sports Physical Therapy*, 14, 70-74.
- DAWSON, B. & TRAPP, R. G. 2001. *Basic and clinical biostatistics*, Singapore, McGraw-Hill.
- DEBI, R., MOR, A., SEGAL, O., SEGAL, G., DEBBI, E., AGAR, G., HALPERIN, N., HAIM, A. & ELBAZ, A. 2009. Differences in gait patterns, pain, function and quality of life between males and females with knee osteoarthritis: a clinical trial. *Bmc Musculoskeletal Disorders*, 10, 127.
- DEHAAN, M. N., GUZMAN, J., BAYLEY, M. T. & BELL, M. J. 2007. Knee osteoarthritis clinical practice guidelines -- how are we doing? *Journal of Rheumatology*, 34, 2099-105.
- DEKKER, J., VAN DIJK, G. M. & VEENHOF, C. 2009. Risk factors for functional decline in osteoarthritis of the hip or knee. *Current Opinion in Rheumatology*, 21, 520-524.
- DERVIN, G. F., STIELL, I. G., WELLS, G. A., RODY, K. & GRABOWSKI, J. 2001. Physicians' accuracy and interrator reliability for the diagnosis of unstable meniscal tears in patients having osteoarthritis of the knee. *Canadian Journal of Surgery*, 44, 267-74.
- DEWALT, D. A., ROTHROCK, N., YOUNT, S., STONE, A. A. & GROUP, P. C. 2007. Evaluation of item candidates: the PROMIS qualitative item review. *Medical Care*, 45, S12-21.
- DEYLE, G. D., ALLISON, S. C., MATEKEL, R. L., RYDER, M. G., STANG, J. M., GOHDES, D. D., HUTTON, J. P., HENDERSON, N. E. & GARBER, M. B. 2005. Physical therapy treatment effectiveness for osteoarthritis of the knee:

- a randomized comparison of supervised clinical exercise and manual therapy procedures versus a home exercise program. *Physical Therapy*, 85, 1301-17.
- DEYLE, G. D., HENDERSON, N. E., MATEKEL, R. L., RYDER, M. G., GARBER, M. B. & ALLISON, S. C. 2000. Effectiveness of manual physical therapy and exercise in osteoarthritis of the knee. A randomized, controlled trial. *Annals of Internal Medicine*, 132, 173-81.
- DIEPPE, P., CUSHNAGHAN, J. & SHEPSTONE, L. 1997. The "Bristol OA 500" study: progression of osteoarthritis (OA) over 3 years and the relationship between clinical and radiographic changes at the knee joint. *Osteoarthritis and Cartilage*, 5, 87-97.
- DIEPPE, P., CUSHNAGHAN, J., YOUNG, P. & KIRWAN, J. 1993. Prediction of the progression of joint space narrowing in osteoarthritis of the knee by bone scintigraphy. *Annals of the Rheumatic Diseases*, 52, 557-563.
- DOUGADOS, M., GUEGUEN, A., NGUYEN, M., THIESCE, A., LISTRAT, V., JACOB, L., NAKACHE, J., GABRIEL, K., LEQUESNE, M. & AMOR, B. 1992. Longitudinal Radiologic Evaluation of Osteoarthritis of the Knee. *Journal of Rheumatology*, 19, 378-84.
- DUNCAN, R., PEAT, G., THOMAS, E., HAY, E., MCCALL, I. & CROFT, P. 2007. Symptoms and radiographic osteoarthritis: not as discordant as they are made out to be? *Annals of the Rheumatic Diseases*, 66, 86-91.
- DUNCAN, R., PEAT, G., THOMAS, E., WOOD, L., HAY, E. & CROFT, P. 2009. Does isolated patellofemoral osteoarthritis matter? *Osteoarthritis & Cartilage*, 17, 1151-5.
- DUNN, J. C. & IVERSEN, M. D. 2003. Hand-held Dynamometer from Elderly Subjects with Degenerative Back Pain. *Journal of Geriatric Physical Therapy*, 26, 23-29.
- DWORKIN, R. H., TURK, D. C., WYRWICH, K. W., BEATON, D., CLEELAND, C. S., FARRAR, J. T., HAYTHORNTHWAITE, J. A., JENSEN, M. P., KERNS, R. D., ADER, D. N., BRANDENBURG, N., BURKE, L. B., CELLA, D., CHANDLER, J., COWAN, P., DIMITROVA, R., DIONNE, R., HERTZ, S., JADAD, A. R., KATZ, N. P., KEHLET, H., KRAMER, L. D., MANNING, D. C., MCCORMICK, C., MCDERMOTT, M. P., MCQUAY, H. J., PATEL, S., PORTER, L., QUESSY, S., RAPPAPORT, B. A., RAUSCHKOLB, C., REVICKI, D. A., ROTHMAN, M., SCHMADER, K. E., STACEY, B. R., STAUFFER, J. W., VON STEIN, T., WHITE, R. E., WITTER, J. & ZAVISIC, S. 2008. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. *Journal of Pain*, 9, 105-21.
- ELVERU, R. A., ROTHSTEIN, J. M. & LAMB, R. L. 1988. Goniometric Reliability in a Clinical Setting: Subtalar and Ankle Joint Measurements. *Physical Therapy*, 68, 672.

- FEINSTEIN, A. R. & CICCHETTI, D. V. 1990. High agreement but low kappa: I. The problems of two paradoxes. *Journal of Clinical Epidemiology*, 43, 543-9.
- FELSON, D. T., GOGGINS, J., NIU, J., ZHANG, Y. & HUNTER, D. J. 2004. The effect of body weight on progression of knee osteoarthritis is dependent on alignment. *Arthritis & Rheumatism*, 50, 3904-9.
- FELSON, D. T. Z., Y. HANNAN, M.T. NAIMARK, A. WEISSMAN, B.N. ALIABADI, P. 1995. The incidence and natural history of knee osteoarthritis in the elderly. *Arthritis & Rheumatism*, 38, 1500 - 1505.
- FIELDING, S., FAYERS, P. M., MCDONALD, A., MCPHERSON, G., CAMPBELL, M. K. & GROUP, R. S. 2008. Simple imputation methods were inadequate for missing not at random (MNAR) quality of life data. *Health & Quality of Life Outcomes*, 6, 57.
- FITZGERALD, G. K. 2010. Invited commentary. *Physical Therapy*, 90, 855-8; author reply 858.
- FITZGERALD, G. K., CHILDS, J. D., RIDGE, T. M. & IRRGANG, J. J. 2002. Agility and perturbation training for a physically active individual with knee osteoarthritis. *Physical Therapy*, 82, 372-82.
- FITZGERALD, G. K. & MCCLURE, P. W. 1995. Reliability of Measurements Obtained With Four Tests for Patellofemoral Alignment. *Physical Therapy*, 75, 84-92.
- FITZGERALD, G. K. & OATIS, C. 2004. Role of physical therapy in management of knee osteoarthritis. *Current Opinion in Rheumatology*, 16, 143-7.
- FITZGERALD, G. K., PIVA, S. R. & IRRGANG, J. J. 2004a. Reports of joint instability in knee osteoarthritis: Its prevalence and relationship to physical function. *Arthritis Care and Research*, 51, 941-6.
- FITZGERALD, G. K., PIVA, S. R., IRRGANG, J. J., BOUZUBAR, F. & STARZ, T. W. 2004b. Quadriceps activation failure as a moderator of the relationship between quadriceps strength and physical function in individuals with knee osteoarthritis. *Arthritis & Rheumatism*, 51, 40-8.
- FORD-SMITH, C. D., WYMAN, J. F., ELSWICK, R. K., JR. & FERNANDEZ, T. 2001. Reliability of stationary dynamometer muscle strength testing in community-dwelling older adults. *Archives of Physical Medicine & Rehabilitation*, 82, 1128-32.
- FOROUGH, N., SMITH, R. M., LANGE, A. K., BAKER, M. K., FIATARONE SINGH, M. A. & VANWANSEELE, B. 2010. Dynamic alignment and its association with knee adduction moment in medial knee osteoarthritis. *Knee*, 17, 210-6.
- FRANSEN, M. 2004. When is physiotherapy appropriate? *Best Practice & Research in Clinical Rheumatology*, 18, 477-89.

- FRANSEN, M., CROSBIE, J. & EDMONDS, J. 2001. Physical therapy is effective for patients with osteoarthritis of the knee: a randomised controlled clinical trial. *The Journal of Rheumatology*, 28, 156-164
- FRANSEN, M. & MCCONNELL, S. 2009. Land-based exercise for osteoarthritis of the knee: a metaanalysis of randomized controlled trials. *Journal of Rheumatology*, 36, 1109-17.
- FRITZ, J. M. 2009. Clinical prediction rules in physical therapy: coming of age? *Journal of Orthopaedic & Sports Physical Therapy*, 39, 159-61.
- FRITZ, J. M., DELITTO, A., ERHARD, R. E. & ROMAN, M. 1998. An examination of the selective tissue tension scheme, with evidence for the concept of a capsular pattern of the knee. *Physical Therapy*, 78, 1046-61.
- FURLAN, A. D., IRVIN, E. & BOMBARDIER, C. 2006. Limited search strategies were effective in finding relevant nonrandomized studies. *Journal of Clinical Epidemiology*, 59, 1303-11.
- FURUKAWA, T. A., STRAUSS, S., BUCHER, H. C. & GUYATT, G. 2008. Diagnostic tests. In: GUYATT, G., RENNIE, D., MEADE, M. O. & COOK, D. J. (eds.) *Users guide to the medical literature: a manual for evidence-based clinical practice*. 2nd ed.: McGraw-Hill Companies.
- GADBURY, G. L., COFFEY, C. S. & ALLISON, D. B. 2003. Modern statistical methods for handling missing repeated measurements in obesity trial data: beyond LOCF. *Obesity Reviews*, 4, 175-84.
- GIBSON, K., SAYERS, S. & MINOR, M. 2010. Measurement of varus/valgus alignment in obese individuals with knee osteoarthritis. *Arthritis Care and Research*, 62, 690-696.
- GILL, S. & MCBURNEY, H. 2008. Reliability of performance-based measures in people awaiting joint replacement surgery of the hip or knee. *Physiotherapy Research International*, 13, 141-52.
- GOLIGHTLY, Y., ALLEN, K., HELMICK, C., SCHWARTZ, T., RENNER, J. & JORDAN, J. 2010. Hazard of incident and progressive knee and hip radiographic osteoarthritis and chronic joint symptoms in individuals with and without limb length inequality. *Journal of Rheumatology*, 37, 2133-2140.
- GOSSEC, L., HAWKER, G., DAVIS, A. M., MAILLEFERT, J. F., LOHMANDER, L. S., ALTMAN, R., CIBERE, J., CONAGHAN, P. G., HOCHBERG, M. C., JORDAN, J. M., KATZ, J. N., MARCH, L., MAHOMED, N., PAVELKA, K., ROOS, E. M., SUAREZ-ALMAZOR, M. E., ZANOLI, G. & DOUGADOS, M. 2007. OMERACT/OARSI initiative to define states of severity and indication for joint replacement in hip and knee osteoarthritis. *Journal of Rheumatology*, 34, 1432-5.
- GURALNIK, J. M., SIMONSICK, E. M., FERRUCCI, L., GLYNN, R. J., BERKMAN, L. F., BLAZER, D. G., SCHERR, P. A. & WALLACE, R. B. 1994. A short physical performance battery assessing lower extremity function:

- association with self-reported disability and prediction of mortality and nursing home admission. *Journal of Gerontology*, 49, M85-94.
- GUYATT, G. H., NORMAN, G. R., JUNIPER, E. F. & GRIFFITH, L. E. 2002. A critical look at transition ratings. *Journal of Clinical Epidemiology*, 55, 900-908.
- HANCOCK, M., HERBERT, R. D. & MAHER, C. G. 2009. A guide to interpretation of studies investigating subgroups of responders to physical therapy interventions. *Physical Therapy*, 89, 698-704.
- HARVEY, W., YANG, M., COOKE, T., SEGAL, N., LANE, N., LEWIS, C. & FELSON, D. 2010. Association of leg-length inequality with knee osteoarthritis a cohort study. *Annals of Internal Medicine*, 152, 287-295.
- HAUKOOS, J. S. & NEWGARD, C. D. 2007. Advanced statistics: missing data in clinical research--part 1: an introduction and conceptual framework. *Academic Emergency Medicine*, 14, 662-8.
- HAWKER, G. A. 2009. Experiencing painful osteoarthritis: what have we learned from listening? *Current Opinion in Rheumatology*, 21, 507-12.
- HAWKER, G. A., DAVIS, A. M., FRENCH, M. R., CIBERE, J., JORDAN, J. M., MARCH, L., SUAREZ-ALMAZOR, M., KATZ, J. N. & DIEPPE, P. 2008a. Development and preliminary psychometric testing of a new OA pain measure--an OARSI/OMERACT initiative. *Osteoarthritis & Cartilage*, 16, 409-14.
- HAWKER, G. A., STEWART, L., FRENCH, M. R., CIBERE, J., JORDAN, J. M., MARCH, L., SUAREZ-ALMAZOR, M. & GOOBERMAN-HILL, R. 2008b. Understanding the pain experience in hip and knee osteoarthritis - an OARSI/OMERACT initiative. *Osteoarthritis and Cartilage*, 16, 415-422.
- HAYDEN, J. A., COTE, P. & BOMBARDIER, C. 2006. Evaluation of the quality of prognosis studies in systematic reviews. *Annals of Internal Medicine*, 144, 427-37.
- HAYES, K. W., PETERSEN, C. & FALCONER, J. 1994. An examination of Cyriax's passive motion tests with patients having osteoarthritis of the knee. *Physical Therapy*, 74, 697-707; discussion 707-9.
- HAYES, K. W. & PETERSEN, C. M. 2001. Reliability of assessing end-feel and pain and resistance sequence in subjects with painful shoulders and knees. *Journal of Orthopaedic & Sports Physical Therapy*, 31, 432-45.
- HEFFORD, C., ABBOTT, J. H., BAXTER, G. D. & ARNOLD, R. 2011. Outcome measurement in clinical practice: practical and theoretical issues for health related quality of life (HRQOL) questionnaires. *Physical Therapy Reviews*, 16, 155-167.
- HELIO LE GRAVERAND, M., BRANDT, K., MAZZUCA, S., RAUNIG, D. & VIGNON, E. 2009. Progressive increase in body mass index is not associated with a progressive increase in joint space narrowing in obese

- women with osteoarthritis of the knee. *Annals of the Rheumatic Diseases*, 68, 1734-1738.
- HERRINGTON, L. C. 2000. The inter-tester reliability of a clinical measurement used to determine the medial/lateral orientation of the patella. *Manual Therapy*, 7, 163-167.
- HICKS, G. E., FRITZ, J. M., DELITTO, A. & MISHOCK, J. 2003. Interrater reliability of clinical examination measures for identification of lumbar segmental instability. *Archives of Physical Medicine & Rehabilitation*, 84, 1858-64.
- HINMAN, R., BOWLES, K. & BENNELL, K. 2009. Laterally wedged insoles in knee osteoarthritis: Do biomechanical effects decline after one month of wear? *Bmc Musculoskeletal Disorders*, 10.
- HINMAN, R. S. & CROSSLEY, K. M. 2007. Patellofemoral joint osteoarthritis: an important subgroup of knee osteoarthritis. *Rheumatology*, 46, 1057-62.
- HINMAN, R. S., HUNT, M. A., CREABY, M. W., WRIGLEY, T. V., MCMANUS, F. J. & BENNELL, K. L. 2010. Hip muscle weakness in individuals with medial knee osteoarthritis. *Arthritis Care & Research*, 62, 1190-3.
- HINMAN, R. S., MAY, R. L. & CROSSLEY, K. M. 2006. Is there an alternative to the full-leg radiograph for determining knee joint alignment in osteoarthritis? *Arthritis & Rheumatism*, 55, 306-13.
- HINMAN, R. S., PAYNE, C., METCALF, B. R., WRIGLEY, T. V. & BENNELL, K. L. 2008. Lateral wedges in knee osteoarthritis: what are their immediate clinical and biomechanical effects and can these predict a three-month clinical outcome? *Arthritis & Rheumatism*, 59, 408-15.
- HORTOBAGYI, T., WESTERKAMP, L., BEAM, S., MOODY, J., GARRY, J., HOLBERT, D. & DEVITA, P. 2005. Altered hamstring-quadriceps muscle balance in patients with knee osteoarthritis. *Clinical Biomechanics*, 20, 97-104.
- HUNTER, D. J. 2009. Focusing osteoarthritis management on modifiable risk factors and future therapeutic prospects. *Therapeutic Advances in Musculoskeletal Disease*, 1, 35-47.
- HUNTER, D. J., LE GRAVERAND, M.-P. H. & ECKSTEIN, F. 2009a. Radiologic markers of osteoarthritis progression. *Current Opinion in Rheumatology*, 21, 110-7.
- HUNTER, D. J. & LO, G. H. 2008. The management of osteoarthritis: an overview and call to appropriate conservative treatment. *Rheumatic Diseases Clinics of North America*, 34, 689-712.
- HUNTER, D. J., SHARMA, L. & SKAIFE, T. 2009b. Alignment and osteoarthritis of the knee. *Journal of Bone & Joint Surgery, American Volume*, 91 Suppl 1, 85-89.
- HUNTER, D. J., ZHANG, Y. Q., NIU, J. B., FELSON, D. T., KWONG, K., NEWMAN, A., KRITCHEVSKY, S., HARRIS, T., CARBONE, L. & NEVITT,

- M. 2007. Patella malalignment, pain and patellofemoral progression: the Health ABC Study. *Osteoarthritis & Cartilage*, 15, 1120-7.
- HURLEY, L., YARDLEY, K., GROSS, A. R., HENDRY, L. & MCLAUGHLIN, L. 2002. A survey to examine attitudes and patterns of practice of physiotherapists who perform cervical spine manipulation. *Manual Therapy*, 7, 10-8.
- HUSSEY, J. & WILSON, F. 2003. Measurement of activity levels is an important part of physiotherapy assessment. *Physiotherapy (London)*, 89, 585-93.
- IRRGANG, J. J., SNYDER-MACKLER, L., WAINNER, R. S., FU, F. H. & HARNER, C. D. 1998. Development of a patient-reported measure of function of the knee: An electromyographic study of vastus medialis oblique and vastus lateralis activity while ascending and descending steps. *Journal of Bone and Joint Surgery (American)*, 80, 1132-45.
- JAESCHKE, R., GUYATT, G. H. & SACKETT, D. L. 1994. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA*, 271, 703-7.
- JAESCHKE, R., SINGER, J. & GUYATT, G. H. 1989. Measurement of health status. Ascertaining the minimal clinically important difference. *Controlled Clinical Trials*, 10, 407-15.
- JAMTVEDT, G., DAHM, K. T., CHRISTIE, A., MOE, R. H., HAAVARDSHOLM, E., HOLM, I. & HAGEN, K. B. 2008a. Physical therapy interventions for patients with osteoarthritis of the knee: an overview of systematic reviews. *Physical Therapy*, 88, 123-36.
- JAMTVEDT, G., DAHM, K. T., HOLM, I. & FLOTTORP, S. 2008b. Measuring physiotherapy performance in patients with osteoarthritis of the knee: a prospective study. *BMC Health Services Research*, 8, 145.
- JANSEN, M. J., VIECHTBAUER, W., LENSSEN, A. F., HENDRIKS, E. J. M. & DE BIE, R. A. 2011. Strength training alone, exercise therapy alone, and exercise therapy with passive manual mobilisation each reduce pain and disability in people with knee osteoarthritis: a systematic review. *Journal of physiotherapy*, 57, 11-20.
- JARIC, S. 2002. Muscle strength testing: use of normalisation for body size. *Sports Medicine*, 32, 615-31.
- JELLEMA, P., VAN DER HORST, H. E., VLAEYEN, J. W. S., STALMAN, W. A. B., BOUTER, L. M. & VAN DER WINDT, D. A. W. M. 2006. Predictors of outcome in patients with (sub)acute low back pain differ across treatment groups. *Spine*, 31, 1699-705.
- JENSEN, M. P., TURNER, J. A., ROMANO, J. M. & FISHER, L. D. 1999. Comparative reliability and validity of chronic pain intensity measures. *Pain*, 83, 157-62.

- JETTE, A. M., JETTE, D. U., NG, J., PLOTKIN, D. J. & BACH, M. A. 1999. Are performance-based measures sufficiently reliable for use in multicenter trials? Musculoskeletal Impairment (MSI) Study Group. *Journals of Gerontology Series A-Biological Sciences & Medical Sciences*, 54, M3-6.
- JINKS, C., JORDAN, K. P., BLAGOJEVIC, M. & CROFT, P. 2008. Predictors of onset and progression of knee pain in adults living in the community. A prospective study. *Rheumatology*, 47, 368-74.
- JONES, A., HOPKINSON, N., PATTRICK, M., BERMAN, P. & DOHERTY, M. 1992. Evaluation of a method for clinically assessing osteoarthritis of the knee. *Annals of the Rheumatic Diseases*, 51, 243-5.
- JONSON, S. R. & GROSS, M. T. 1997. Intraexaminer Reliability, Interexaminer Reliability, and Mean values for Nine Lower Extremity Skeletal Measures in Healthy Naval Midshipmen. *Journal of Orthopaedic & Sports Physical Therapy*, 25, 253-263.
- JORDAN, K., JINKS, C. & CROFT, P. 2006. A prospective study of the consulting behaviour of older people with knee pain. *British Journal of General Practice*, 56, 269-76.
- KAMPER, S. 2009. Global rating of change scales. *Australian Journal of Physiotherapy*, 55, 289.
- KEEFE, F. J., LEFEBVRE, J. C., EGERT, J. R., AFFLECK, G., SULLIVAN, M. J. & CALDWELL, D. S. 2000. The relationship of gender to pain, pain behavior, and disability in osteoarthritis patients: the role of catastrophizing. *Pain*, 87, 325-34.
- KENNEDY, D. M., STRATFORD, P. W., WESSEL, J., GOLLISH, J. D. & PENNEY, D. 2005. Assessing stability and change of four performance measures: a longitudinal study evaluating outcome following total hip and knee arthroplasty. *Bmc Musculoskeletal Disorders*, 6, 3.
- KIRKWOOD, B. R. & STERNE, A. C. 2003. *Essential medical statistics*, Victoria, Blackwell Science Asia Pty Ltd.
- KRAEMER, H. C., WILSON, G. T., FAIRBURN, C. G. & AGRAS, W. S. 2002. Mediators and moderators of treatment effects in randomized clinical trials. *Archives of General Psychiatry*, 59, 877-83.
- KRAUS, V. B. 2006. Do biochemical markers have a role in osteoarthritis diagnosis and treatment? *Best Practice & Research in Clinical Rheumatology*, 20, 69-80.
- KVIEN, T. K., HEIBERG, T. & HAGEN, K. B. 2007. Minimal clinically important improvement/difference (MCII/MCID) and patient acceptable symptom state (PASS): what do these concepts mean? *Annals of the Rheumatic Diseases*, 66 Suppl 3, iii40-1.
- LANDIS, J. R. & KOCH, G. G. 1977. The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.

- LANTZ, C. A. 1997. Application and evaluation of the kappa statistic in the design and interpretation of chiropractic clinical research. *Journal of Manipulative & Physiological Therapeutics*, 20, 521-8.
- LAUPACIS, A., SEKAR, N. & STIELL, I. G. 1997. Clinical prediction rules. A review and suggested modifications of methodological standards. *JAMA*, 277, 488-494.
- LEDINGHAM, J., REGAN, M., JONES, A. & DOHERTY, M. 1995. Factors Affecting Radiographic Progression of Knee Osteoarthritis. *Annals of the Rheumatic Diseases*, 54, 53-58.
- LESHER, J. D., SUTLIVE, T. G., MILLER, G. A., CHINE, N. J., GARBER, M. B. & WAINNER, R. S. 2006. Development of a clinical prediction rule for classifying patients with patellofemoral pain syndrome who respond to patellar taping. *Journal of Orthopaedic & Sports Physical Therapy*, 36, 854-66.
- LIEVENSE, A. M., BIERMA-ZEINSTR, S. M. A., VERHAGEN, A. P., VERHAAR, J. A. N. & KOES, B. W. 2002. Prognostic factors of progress of hip osteoarthritis: a systematic review. *Arthritis & Rheumatism*, 47, 556-62.
- LIM, B.-W., HINMAN, R. S., WRIGLEY, T. V., SHARMA, L. & BENNELL, K. L. 2008. 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 & Rheumatism*, 59, 943-51.
- LIN, Y. C., DAVEY, R. C. & COCHRANE, T. 2001. Tests for physical function of the elderly with knee and hip osteoarthritis. *Scandinavian Journal of Medicine & Science in Sports*, 11, 280-6.
- LIU, G. & GOULD, A. L. 2002. Comparison of alternative strategies for analysis of longitudinal trials with dropouts. *Journal of Biopharmaceutical Statistics*, 12, 207-26.
- MAGEE, D. J. 2002. *Orthopaedic physical examination*, Philadelphia, Saunders Elsevier.
- MAITLAND, G. D. 1986. *Vertebral Manipulation*, Oxford, Butterworth Heinemann.
- MAITLAND, G. D. 1991. *Peripheral Manipulation*, Oxford, Butterworth Heinemann.
- MALANGA, G. A. & NADLER, S. F. 2006. *Musculoskeletal physical examination. An evidence based approach.*, Philadelphia, Mosby Elsevier.
- MALLEN, C., PEAT, G., PORCHERET, M. & CROFT, P. 2007a. The prognosis of joint pain in the older patient: General practitioners' views on discussing and estimating prognosis. *European Journal of General Practice*, 13, 166-168.
- MALLEN, C. D. & PEAT, G. 2009. Discussing prognosis with older people with musculoskeletal pain: a cross-sectional study in general practice. *BMC family practice*, 10, 50.

- MALLEN, C. D., PEAT, G., THOMAS, E., LACEY, R. & CROFT, P. 2007b. Predicting poor functional outcome in community-dwelling older adults with knee pain: prognostic value of generic indicators. *Annals of the Rheumatic Diseases*, 66, 1456-61.
- MANCHIKANTI, L., DATTA, S., SMITH, H. S. & HIRSCH, J. A. 2009. Evidence-based medicine, systematic reviews, and guidelines in interventional pain management: part 6. Systematic reviews and meta-analyses of observational studies. *Pain Physician*, 12, 819-50.
- MANNION, A. F., PORCHET, F., KLEINSTUCK, F. S., LATTIG, F., JESZENSZKY, D., BARTANUSZ, V., DVORAK, J. & D., G. 2009. The quality of spine surgery from the patient's perspective: Part 2. Minimal clinically important difference for improvement and deterioration as measured with the Core Outcome Measures Index. *European Spine Journal*, 18, S374-S379.
- MAWDSLEY, R., MORAN, K. & CONNIFF, L. 2002. Reliability of two commonly used pain scales with elderly patients. *Journal of Geriatric Physical Therapy*, 25, 16-20.
- MAZZUCA, S. A., BRANDT, K. D., KATZ, B. P., DING, Y., LANE, K. A., BUCKWALTER, K. A., MAZZUCA, S. A., BRANDT, K. D., KATZ, B. P., DING, Y., LANE, K. A. & BUCKWALTER, K. A. 2006. Risk factors for progression of tibiofemoral osteoarthritis: an analysis based on fluoroscopically standardised knee radiography. *Annals of the Rheumatic Diseases*, 65, 515-9.
- MAZZUCA, S. A., BRANDT, K. D., SCHAUWECKER, D. S., KATZ, B. P., MEYER, J. M., LANE, K. A., BRADLEY, J. D., HUGENBERG, S. T., WOLFE, F., MORELAND, L. W., HECK, L. W., YOCUM, D. E., SCHNITZER, T. J., SHARMA, L., MANZI, S., ODDIS, C. V., MAZZUCA, S. A., BRANDT, K. D., SCHAUWECKER, D. S., KATZ, B. P., MEYER, J. M., LANE, K. A., BRADLEY, J. D., HUGENBERG, S. T., WOLFE, F., MORELAND, L. W., HECK, L. W., YOCUM, D. E., SCHNITZER, T. J., SHARMA, L., MANZI, S. & ODDIS, C. V. 2005. Severity of joint pain and Kellgren-Lawrence grade at baseline are better predictors of joint space narrowing than bone scintigraphy in obese women with knee osteoarthritis. *Journal of Rheumatology*, 32, 1540-6.
- MCALINDON, T. E., FELSON, D. T., ZHANG, Y., HANNAN, M. T., ALIABADI, P., WEISSMAN, B., RUSH, D., WILSON, P. W. F. & JACQUES, P. 1996a. Relation of Dietary Intake and Serum Levels Of Vitamin D to Progression of Osteoarthritis of the Knee among Participants in the Framingham Study. *Annals of Internal Medicine*, 125, 353-359.
- MCALINDON, T. E., JACQUES, P., ZHANG, Y., HANNAN, M. T., ALIABADI, P., WEISSMAN, B., RUSH, D., LEVY, D. & FELSON, D. T. 1996b. Do

- antioxidant micronutrients protect against the development and progression of knee osteoarthritis? *Arthritis & Rheumatism*, 39, 648-656.
- MCEWAN, I., HERRINGTON, L. & THOM, J. 2007. The validity of clinical measures of patella position. *Manual Therapy*, 12, 226-230.
- MCGINN, T. G., GUYATT, G. H., WYER, P. C., NAYLOR, C. D., STIELL, I. G. & RICHARDSON, W. S. 2000. Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. Evidence-Based Medicine Working Group. *JAMA*, 284, 79-84.
- MCPOIL, T. G. & CORNWALL, M. W. 2005. Use of the Longitudinal Arch Angle to Predict Dynamic Foot Posture in Walking. *Journal of American Podiatry Medical Association* 95, 114-120.
- MIKESKY, A. E., MAZZUCA, S. A., BRANDT, K. D., PERKINS, S. M., DAMUSH, T. & LANE, K. A. 2006. Effects of strength training on the incidence and progression of knee osteoarthritis. *Arthritis & Rheumatism*, 55, 690-9.
- MITCHELL, H. L., CARR, A. J. & SCOTT, D. L. 2006. The management of knee pain in primary care: factors associated with consulting the GP and referrals to secondary care. *Rheumatology*, 45, 771-6.
- MOHER, D., LIBERATI, A., TETZLAFF, J. & ALTMAN, D. G. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *British Medical Journal*, 339, 332-336.
- MOKKINK, L. B., TERWEE, C. B., VAN DER SLIKKE, R. M. A., VAN LUMMEL, R. C., BENINK, R. J., BOUTER, L. M. & DE VET, H. C. W. 2005. Reproducibility and validity of the DynaPort KneeTest. *Arthritis & Rheumatism*, 53, 357-63.
- MOLNAR, F. J., HUTTON, B. & FERGUSON, D. 2008. Does analysis using "last observation carried forward" introduce bias in dementia research? *CMAJ Canadian Medical Association Journal*, 179, 751-3.
- MOONS, K. G. M., ALTMAN, D. G., VERGOUWE, Y. & ROYSTON, P. 2009a. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ*, 338, b606.
- MOONS, K. G. M., ROYSTON, P., VERGOUWE, Y., GROBBEE, D. E. & ALTMAN, D. G. 2009b. Prognosis and prognostic research: what, why, and how? *BMJ*, 338, b375.
- MOSKOWITZ, R. W., ALTMAN, R. D., HOCHBERG, M. C., BUCKWALTER, J. A. & GOLDBERG, V. M. (eds.) 2007. *Osteoarthritis. Diagnosis and medical/surgical management.*, Philadelphia: Lippincott Williams and Wilkins.
- MURPHY, L., SCHWARTZ, T. A., HELMICK, C. G., RENNER, J. B., TUDOR, G., KOCH, G., DRAGOMIR, A., KALSBECK, W. D., LUTA, G. & JORDAN, J.

- M. 2008. Lifetime risk of symptomatic knee osteoarthritis. *Arthritis & Rheumatism*, 59, 1207-13.
- NATIONAL COLLABORATING CENTRE FOR CHRONIC CONDITIONS 2008. Osteoarthritis: national clinical guideline for care and management in adults. . London: Royal College of Physicians.
- NEOGI, T., NEVITT, M., YANG, M., CURTIS, J., TORNER, J. & FELSON, D. 2010. Consistency of knee pain: Correlates and association with function. *Osteoarthritis and Cartilage*, 18, 1250-1255.
- NEWGARD, C. D. & HAUKOOS, J. S. 2007. Advanced statistics: missing data in clinical research--part 2: multiple imputation. *Academic Emergency Medicine*, 14, 669-78.
- NIU, J., ZHANG, Y., TORNER, J., NEVITT, M., LEWIS, C., ALIABADI, P., SACK, B., CLANCY, M., SHARMA, L. & FELSON, D. 2009. Is obesity a risk factor for progressive radiographic knee osteoarthritis? *Arthritis Care and Research*, 61, 329-335.
- NUNNALLY, J. C. 1978. *Psychometric Theory*, New York, McGraw-Hill.
- OTTAWA PANEL 2005. Ottawa panel evidence-based clinical practice guidelines for therapeutic exercises and manual therapy in the management of osteoarthritis. *Physical Therapy*, 85, 907-71.
- PANNELL, D. J. 1997. Sensitivity analysis of normative economic models: theoretical framework and practical strategies. *Agricultural Economics*, 16, 139-152.
- PEAT, G., MCCARNEY, R. & CROFT, P. 2001. Knee pain and osteoarthritis in older adults: a review of community burden and current use of primary health care. *Annals of the Rheumatic Diseases*, 60, 91-7.
- PEAT, G., THOMAS, E., DUNCAN, R. & WOOD, L. 2010. Is a "false-positive" clinical diagnosis of knee osteoarthritis just the early diagnosis of pre-radiographic disease? *Arthritis Care & Research*, 62, 1502-6.
- PEAT, G., WOOD, L., WILKIE, R., THOMAS, E. & GROUP, K.-S. S. 2003. How reliable is structured clinical history-taking in older adults with knee problems? Inter- and intraobserver variability of the KNE-SCI. *Journal of Clinical Epidemiology*, 56, 1030-7.
- PETERS, T. J., SANDERS, C., DIEPPE, P., DONOVAN, J., PETERS, T. J., SANDERS, C., DIEPPE, P., & DONOVAN, J. 2005. Factors associated with change in pain and disability over time: a community-based prospective observational study of hip and knee osteoarthritis. *British Journal of General Practice*, 55, 205-11.
- PETRIE, A. 2006. Statistics in orthopaedic papers. *Journal of Bone & Joint Surgery - British Volume*, 88-B, 1121-1136.
- PETTERSON, S. C., BARRANCE, P., BUCHANAN, T., BINDER-MACLEOD, S. & SNYDER-MACKLER, L. 2008. Mechanisms underlying quadriceps

- weakness in knee osteoarthritis. *Medicine & Science in Sports & Exercise*, 40, 422-7.
- PHAM, T., VAN DER HEIJDE, D., ALTMAN, R. D., ANDERSON, J. J., BELLAMY, N., HOCHBERG, M., SIMON, L., STRAND, V., WOODWORTH, T. & DOUGADOS, M. 2004. OMERACT-OARSI initiative: Osteoarthritis Research Society International set of responder criteria for osteoarthritis clinical trials revisited. *Osteoarthritis & Cartilage*, 12, 389-99.
- PILLING, S., ANDERSON, I., GOLDBERG, D., MEADER, N., TAYLOR, C. & TWO GUIDELINE DEVELOPMENT, G. 2009. Depression in adults, including those with a chronic physical health problem: summary of NICE guidance. *BMJ*, 339, b4108.
- PIVA, S. R., FITZGERALD, G. K., IRRGANG, J. J., BOUZUBAR, F. & STARZ, T. W. 2004. Get up and go test in patients with knee osteoarthritis. *Archives of Physical Medicine & Rehabilitation*, 85, 284-9.
- PODSIADLO, D. & RICHARDSON, S. 1991. The timed "Up & Go": a test of basic functional mobility for frail elderly persons. *Journal of the American Geriatrics Society*, 39, 142-8.
- POITRAS, S., ROSSIGNOL, M., AVOUAC, J., AVOUAC, B., CEDRASCHI, C., NORDIN, M., ROUSSEAUX, C., ROZENBERG, S., SAVARIEAU, B., THOUMIE, P., VALAT, J.-P., VIGNON, E. & HILLIQUIN, P. 2010. Management recommendations for knee osteoarthritis: how usable are they? *Joint, Bone, Spine: Revue du Rhumatisme*, 77, 458-65.
- PORCHERET, M., JORDAN, K., JINKS, C., CROFT, P. & PRIMARY CARE RHEUMATOLOGY, S. 2007. Primary care treatment of knee pain--a survey in older adults. *Rheumatology*, 46, 1694-700.
- PORTNEY, L. G. & WATKINS, M. P. 2000. *Foundations of clinical research: applications to practice*, Upper Saddle River, New Jersey, Prentice Hall Health.
- POTTER, M., GORDON, S. & HAMER, P. 2004. The Nominal Group Technique: A useful consensus methodology in physiotherapy research. *New Zealand Journal of Physiotherapy*, 32, 126-30.
- PRAKASH, A., RISSER, R. C. & MALLINCKRODT, C. H. 2008. The impact of analytic method on interpretation of outcomes in longitudinal clinical trials. *International Journal of Clinical Practice*, 62, 1147-58.
- REFSHAUGE, K. M. & LATIMER, J. 1995. The History. . In: REFSHAUGE, K. M. & GASS, E. (eds.) *Musculoskeletal Medicine: Clinical Science and Practice*. . Oxford: Butterworth Heinemann.
- REID, D. A. & MCNAIR, P. J. 2010. Effects of an acute hamstring stretch in people with and without osteoarthritis of the knee. *Physiotherapy*, 96, 14-21.
- REIJMAN, M., POLS, H. A., BERGINK, A. P., HAZES, J. M., BELO, J. N., LIEVENSE, A. M., BIERMA-ZEINSTRAS, S. M., REIJMAN, M., POLS, H. A.

- P., BERGINK, A. P., HAZES, J. M. W., BELO, J. N., LIEVENSE, A. M. & BIERMA-ZEINSTRAS, S. M. A. 2007. Body mass index associated with onset and progression of osteoarthritis of the knee but not of the hip: the Rotterdam Study.[see comment]. *Annals of the Rheumatic Diseases*, 66, 158-62.
- REILLY, K., BARKER, K., SHAMLEY, D., NEWMAN, M., OSKROCHI, G. & SANDALL, S. 2009. The role of foot and ankle assessment of patients with lower limb osteoarthritis. *Physiotherapy*, 95, 164-169.
- RODDY, E., ZHANG, W., DOHERTY, M., ARDEN, N. K., BARLOW, J., BIRRELL, F., CARR, A., CHAKRAVARTY, K., DICKSON, J., HAY, E., HOSIE, G., HURLEY, M., JORDAN, K. M., MCCARTHY, C., MCMURDO, M., MOCKETT, S., O'REILLY, S., PEAT, G., PENDLETON, A. & RICHARDS, S. 2005. Evidence-based recommendations for the role of exercise in the management of osteoarthritis of the hip or knee--the MOVE consensus. *Rheumatology*, 44, 67-73.
- ROSEMANN, T., KUEHLEIN, T., LAUX, G. & SZECSENYI, J. 2008. Factors associated with physical activity of patients with osteoarthritis of the lower limb. *Journal of Evaluation in Clinical Practice*, 14, 288-93.
- ROY, M.-A. G. & DOHERTY, T. J. 2004. Reliability of hand-held dynamometry in assessment of knee extensor strength after hip fracture. *American Journal of Physical Medicine & Rehabilitation*, 83, 813-8.
- ROYSTON, P., ALTMAN, D. G. & SAUERBREI, W. 2006. Dichotomizing continuous predictors in multiple regression: a bad idea. *Statistics in Medicine*, 25, 127-141.
- ROYSTON, P., MOONS, K. G. M., ALTMAN, D. G. & VERGOUWE, Y. 2009. Prognosis and prognostic research: Developing a prognostic model. *BMJ*, 338, b604.
- SANDBORGH, M., LINDBERG, P. & DENISON, E. 2007. Pain belief screening instrument: Development and preliminary validation of a screening instrument for disabling persistent pain. *Journal of Rehabilitation Medicine*, 39, 461-6.
- SANDBORGH, M., LINDBERG, P. & DENISON, E. 2008. The Pain Belief Screening Instrument (PBSI): predictive validity for disability status in persistent musculoskeletal pain. *Disability & Rehabilitation*, 30, 1123-30.
- SCHMITT, J. & DI FABIO, R. P. 2005. The Validity of Prospective and Retrospective Global Change Criterion Measures. *Archives of Physical Medicine & Rehabilitation*, 86, 2270-2276.
- SCHMITT, L. C., FITZGERALD, G. K., REISMAN, A. S. & RUDOLPH, K. S. 2008. Instability, laxity, and physical function in patients with medial knee osteoarthritis. *Physical Therapy*, 88, 1506-16.

- SCHMITT, L. C. & RUDOLPH, K. S. 2008. Muscle stabilization strategies in people with medial knee osteoarthritis: the effect of instability. *Journal of Orthopaedic Research*, 26, 1180-5.
- SCHOUTEN, J. S., VAN DEN OUWELAND, F. A. & VALKENBURG, H. A. 1992. A 12-year follow-up study in the general population on prognostic factors of cartilage loss in osteoarthritis of the knee. *Annals of the Rheumatic Diseases*, 1992, 932-937.
- SCOPAZ, K., PIVA, S., GIL, A., WOOLLARD, J., ODDIS, C. & FITZGERALD, G. 2009. Effect of baseline quadriceps activation on changes in quadriceps strength after exercise therapy in subjects with knee osteoarthritis. *Arthritis Care and Research*, 61, 951-957.
- SEROR, R., TUBACH, F., BARON, G., FALISSARD, B., LOGEART, I., DOUGADOS, M. & RAVAUD, P. 2008. Individualising the Western Ontario and McMaster Universities osteoarthritis index (WOMAC) function subscale: incorporating patient priorities for improvement to measure functional impairment in hip or knee osteoarthritis. *Annals of Rheumatic Diseases*, 67, 494-499.
- SHARMA, L. 2003. Examination of exercise effects on knee osteoarthritis outcomes: why should the local mechanical environment be considered? *Arthritis & Rheumatism*, 49, 255-60.
- SHARMA, L., CAHUE, S., SONG, J., HAYES, K., PAI, Y. C. & DUNLOP, D. 2003a. Physical functioning over three years in knee osteoarthritis - Role of psychosocial, local mechanical, and neuromuscular factors. *Arthritis and Rheumatism*, 48, 3359-3370.
- SHARMA, L., DUNLOP, D. D., CAHUE, S., SONG, J., HAYES, K. W., SHARMA, L., DUNLOP, D. D., CAHUE, S., SONG, J. & HAYES, K. W. 2003b. Quadriceps strength and osteoarthritis progression in malaligned and lax knees.[see comment][summary for patients in Ann Intern Med. 2003 Apr 15;138(8):I1; PMID: 12693914]. *Annals of Internal Medicine*, 138, 613-9.
- SHARMA, L., ECKSTEIN, F., SONG, J., GUERMAZI, A., PRASAD, P., KAPOOR, D., CAHUE, S., MARSHALL, M., HUDELMAIER, M. & DUNLOP, D. 2008. Relationship of meniscal damage, meniscal extrusion, malalignment, and joint laxity to subsequent cartilage loss in osteoarthritic knees. *Arthritis & Rheumatism*, 58, 1716-26.
- SHARMA, L., SONG, J., DUNLOP, D., FELSON, D., LEWIS, C., SEGAL, N., TORNER, J., COOKE, T., HIETPAS, J., LYNCH, J. & NEVITT, M. 2010. Varus and valgus alignment and incident and progressive knee osteoarthritis. *Annals of the Rheumatic Diseases*, 69, 1940-1945.
- SHARMA, L., SONG, J., FELSON, D. T., CAHUE, S., SHAMIYEH, E., DUNLOP, D. D., SHARMA, L., SONG, J., FELSON, D. T., CAHUE, S., SHAMIYEH, E. & DUNLOP, D. D. 2001. The role of knee alignment in disease progression

- and functional decline in knee osteoarthritis.[erratum appears in JAMA 2001 Aug 15;286(7):792]. *JAMA*, 286, 188-95.
- SHIOZAKI, H., KOGA, Y., OMORI, G. & TAMAKI, M. 1999. Obesity and osteoarthritis of the knee in women: Results from the Matsudai Knee Osteoarthritis Survey. *The Knee*, 6, 189-192.
- SIM, J. & WRIGHT, C. C. 2005. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Physical Therapy*, 85, 257-268.
- SIMUNOVIC, N., SPRAGUE, S. & BHANDARI, M. 2009. Methodological issues in systematic reviews and meta-analyses of observational studies in orthopaedic research. *Journal of Bone & Joint Surgery - American Volume*, 91 Suppl 3, 87-94.
- SMART, K. & DOODY, C. 2007. The clinical reasoning of pain by experienced musculoskeletal physiotherapists. *Manual Therapy*, 12, 40-9.
- SOMERS, T. J., KEEFE, F. J., GODIWALA, N. & HOYLER, G. H. 2009. Psychosocial factors and the pain experience of osteoarthritis patients: new findings and new directions. *Current Opinion in Rheumatology*, 21, 501-6.
- SPECTOR, T. D., DACRE, J. E., HARRIS, P. A. & HUSKISSON, E. C. 1992. Radiological progression of osteoarthritis: An 11 year follow up study of the knee. *Annals of the Rheumatic Diseases*, 51, 1107-1110.
- SPECTOR, T. D., HART, D. J. & DOYLE, D. V. 1994. Incidence and progression of osteoarthritis in women with unilateral knee disease in the general population: The effect of obesity. *Annals of the Rheumatic Diseases*, 53, 565-568.
- SRIKANTH, V. K., FRYER, J. L., ZHAI, G., WINZENBERG, T. M., HOSMER, D. & JONES, G. 2005. A meta-analysis of sex differences prevalence, incidence and severity of osteoarthritis. *Osteoarthritis & Cartilage*, 13, 769-81.
- STANTON, T. R., HANCOCK, M. J., MAHER, C. G. & KOES, B. W. 2010. Critical appraisal of clinical prediction rules that aim to optimize treatment selection for musculoskeletal conditions. *Physical Therapy*, 90, 843-54.
- STATISTICS NEW ZEALAND. 2006. *Quick stats about Dunedin city* [Online]. New Zealand Government. [Accessed].
- STRATFORD, P. W. & GOLDSMITH, C. H. 1997. Use of the standard error as a reliability index of interest: an applied example using elbow flexor strength data. *Physical Therapy*, 77, 745-750.
- STRATFORD, P. W., KENNEDY, D., PAGURA, S. M. C. & GOLLISH, J. D. 2003. The relationship between self-report and performance-related measures: questioning the content validity of timed tests. *Arthritis & Rheumatism*, 49, 535-40.
- STRATFORD, P. W. & KENNEDY, D. M. 2004. Does parallel item content on WOMAC's Pain and Function Subscales limit its ability to detect change in functional status? *BMC Musculoskeletal Disorders*, 5.

- STRATFORD, P. W., KENNEDY, D. M. & WOODHOUSE, L. J. 2006. Performance measures provide assessments of pain and function in people with advanced osteoarthritis of the hip or knee. *Physical Therapy*, 86, 1489-96.
- STREINER, D. L. 1995. Learning how to differ: agreement and reliability statistics in psychiatry. *Canadian Journal of Psychiatry - Revue Canadienne de Psychiatrie*, 40, 60-6.
- STREINER, D. L. 2010. Measure for measure: new developments in measurement and item response theory. *Canadian Journal of Psychiatry - Revue Canadienne de Psychiatrie*, 55, 180-6.
- STREINER, D. L. & NORMAN, G. R. 2003. *Health measurement scales: a practical guide to their development and use*, New York, Oxford University Press.
- SUTLIVE, T. G., MITCHELL, S. D., MAXFIELD, S. N., MCLEAN, C. L., NEUMANN, J. C., SWIECKI, C. R., HALL, R. C., BARE, A. C. & FLYNN, T. W. 2004. Identification of Individuals With Patellofemoral Pain Whose Symptoms Improved After a Combined Program of Foot Orthosis Use and Modified Activity: A Preliminary Investigation. *Physical Therapy*, 84, 49-61.
- TANAMAS, S., HANNA, F., CICUTTINI, F., WLUKA, A., BERRY, P. & URQUHART, D. 2009. Does knee malalignment increase the risk of development and progression of knee osteoarthritis? A systematic review. *Arthritis Care and Research*, 61, 459-467.
- TERWEE, C. B., MOKKINK, L. B., STEULTJENS, M. P. M. & DEKKER, J. 2006. Performance-based methods for measuring the physical function of patients with osteoarthritis of the hip or knee: a systematic review of measurement properties. *Rheumatology*, 45, 890-902.
- TERWEE, C. B., ROORDA, L. D., DEKKER, J., BIERMA-ZEINSTRAS, S. M., PEAT, G., JORDAN, K. P., CROFT, P. & DE VET, H. C. W. 2010. Mind the MIC: large variation among populations and methods. *Journal of Clinical Epidemiology*, 63, 524-34.
- THOMAS, E., PEAT, G., MALLEEN, C., WOOD, L., LACEY, R., DUNCAN, R. & CROFT, P. 2008. Predicting the course of functional limitation among older adults with knee pain: do local signs, symptoms and radiographs add anything to general indicators? *Annals of the Rheumatic Diseases*, 67, 1390-8
- THOMPSON, L. R., BOUDREAU, R., HANNON, M. J., NEWMAN, A. B., CHU, C. R., JANSEN, M., NEVITT, M. C. & KWOH, C. K. 2009. The knee pain map: Reliability of a method to identify knee pain location and pattern. *Arthritis & Rheumatism: Arthritis Care & Research*, 61, 725-731.
- THORSTENSSON, C., ANDERSSON, M., JONSSON, H., SAXNE, T. & PETERSSON, I. 2009a. Natural course of knee osteoarthritis in middle-aged subjects with knee pain: 12-Year follow-up using clinical and radiographic criteria. *Annals of the Rheumatic Diseases*, 68, 1890-1893.

- THORSTENSSON, C. A., GOOBERMAN-HILL, R., ADAMSON, J., WILLIAMS, S. & DIEPPE, P. 2009b. Help-seeking behaviour among people living with chronic hip or knee pain in the community. *Bmc Musculoskeletal Disorders*, 10, 153.
- THORSTENSSON, C. A., PETERSSON, I. F., JACOBSSON, L. T. H., BOEGARD, T. L. & ROOS, E. M. 2004. Reduced functional performance in the lower extremity predicted radiographic knee osteoarthritis five years later. *Annals of the Rheumatic Diseases*, 63, 402-407.
- TOOTH, L. R. & OTTENBACHER, K. J. 2004. The kappa statistic in rehabilitation research: an examination. *Archives of Physical Medicine and Rehabilitation*, 85, 1371-1376.
- TUBACH, F., RAVAUD, P., BARON, G., FALISSARD, B., LOGEART, I., BELLAMY, N., BOMBARDIER, C., FELSON, D., HOCHBERG, M., VAN DER HEIJDE, D. & DOUGADOS, M. 2005. Evaluation of clinically relevant changes in patient reported outcomes in knee and hip osteoarthritis: the minimal clinically important improvement. *Annals of the Rheumatic Diseases*, 64, 29-33.
- UNNEBRINK, K. & WINDELER, J. 2001. Intention-to-treat: methods for dealing with missing values in clinical trials of progressively deteriorating diseases. *Statistics in Medicine*, 20, 3931-46.
- VAN DEN DIKKENBERG, N., MEIJER, O. G., VAN DER SLIKKE, R. M. A., VAN LUMMEL, R. C., VAN DIEEN, J. H., PIJLS, B., BENINK, R. J. & WUISMAN, P. I. J. M. 2002. Measuring functional abilities of patients with knee problems: rationale and construction of the DynaPort knee test. *Knee Surgery, Sports Traumatology, Arthroscopy*, 10, 204-12.
- VAN DER ESCH, M., STEULTJENS, M., KNOL, D. L., DINANT, H. & DEKKER, J. 2006. Joint laxity and the relationship between muscle strength and functional ability in patients with osteoarthritis of the knee. *Arthritis & Rheumatism*, 55, 953-9.
- VAN DER ESCH, M., STEULTJENS, M., WIERINGA, H., DINANT, H. & DEKKER, J. 2005. Structural joint changes, malalignment, and laxity in osteoarthritis of the knee. *Scandinavian Journal of Rheumatology*, 34, 298-301.
- VAN DER WAAL, J. M., BOT, S. D. M., TERWEE, C. B., VAN DER WINDT, D. A. W. M., SCHOLTEN, R. J. P. M., BOUTER, L. M. & DEKKER, J. 2005. Course and prognosis of knee complaints in general practice. *Arthritis & Rheumatism*, 53, 920-30.
- VAN DIJK, G. M., DEKKER, J., VEENHOF, C., VAN DEN ENDE, C. H. M. & CARPA STUDY, G. 2006. Course of functional status and pain in osteoarthritis of the hip or knee: a systematic review of the literature. *Arthritis & Rheumatism*, 55, 779-85.

- VINCENZINO, B., COLLINS, N., CLELAND, J. A. & MCPOIL, T. G. 2010. A clinical prediction rule for identifying patients with patellofemoral pain who are likely to benefit from foot orthoses: a preliminary determination. *British Journal of Sports Medicine*, 44, 862-866.
- WALTER, S. D., ELIASZIW, M. & DONNER, A. 1998. Sample size and optimal designs for reliability studies. *Statistics in Medicine*, 17, 101-110.
- WALTON, M. K. 2009. Addressing and advancing the problem of missing data. *Journal of Biopharmaceutical Statistics*, 19, 945-56.
- WAND, B. M., MCAULEY, J. H., MARSTON, L. & DE SOUZA, L. H. 2009. Predicting Outcome in Acute Low Back Pain Using Different Models of Patient Profiling. *Spine*, 34, 1970-1975.
- WASHBURN, R. A., SMITH, K. W., JETTE, A. M. & JANNEY, C. A. 1993. The Physical Activity Scale for the Elderly (PASE): development and evaluation. *Journal of Clinical Epidemiology*, 46, 153-62.
- WATKINS, G., RIDDLE, D. L., LAMB, R. L. & PERSONIUS, W. J. 1991. Reliability of Goniometric Measurements and Visual Estimates of Knee Range of Motion Obtained in a Clinical Setting. *Physical Therapy*, 71, 90-97.
- WEIGL, M., ANGST, F., AESCHLIMANN, A., LEHMANN, S. & STUCKI, G. 2006. 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*, 14, 641-51.
- WEST, S., KING, V., CAREY, T. S., LOHR, K. N., MCKOY, N., SUTTON, S. F. & LUX, L. 2002. Systems to Rate the Strength Of Scientific Evidence. *Evidence Report/Technology Assessment*. Research Triangle Institute–University of North Carolina Evidence-based Practice Center.
- WHOOLEY, M. A., AVINS, A. L., MIRANDA, J. & BROWNER, W. S. 1997. Case-finding instruments for depression. Two questions are as good as many. *Journal of General Internal Medicine*, 12, 439-45.
- WILCOX, S., BRENES, G. A., LEVINE, D., SEVICK, M. A., SHUMAKER, S. A. & CRAVEN, T. 2000. Factors related to sleep disturbance in older adults experiencing knee pain or knee pain with radiographic evidence of knee osteoarthritis. *Journal of the American Geriatrics Society*, 48, 1241-51.
- WILCZYNSKI, N. L. & HAYNES, R. B. 2005. Optimal search strategies for detecting clinically sound prognostic studies in EMBASE: an analytic survey. *Journal of the American Medical Informatics Association*, 12, 481-5.
- WILCZYNSKI, N. L., HAYNES, R. B. & HEDGES, T. 2004. Developing optimal search strategies for detecting clinically sound prognostic studies in MEDLINE: an analytic survey. *BMC Medicine*, 2, 23.
- WOLFE, F., LANE, N. E., WOLFE, F. & LANE, N. E. 2002. The longterm outcome of osteoarthritis: rates and predictors of joint space narrowing in

- symptomatic patients with knee osteoarthritis. *Journal of Rheumatology*, 29, 139-46.
- WOOD, L., PEAT, G., WILKIE, R., HAY, E., THOMAS, E. & SIM, J. 2006. A study of the noninstrumented physical examination of the knee found high observer variability. *Journal of Clinical Epidemiology*, 59, 512-20.
- WOOD, L. R. J., PEAT, G., THOMAS, E. & DUNCAN, R. 2007. Knee osteoarthritis in community-dwelling older adults: are there characteristic patterns of pain location? *Osteoarthritis & Cartilage*, 15, 615-23.
- WOOLHEAD, G., GOOBERMAN-HILL, R., DIEPPE, P. & HAWKER, G. 2010. Night pain in hip and knee osteoarthritis: a focus group study. *Arthritis Care & Research*, 62, 944-9.
- WORLD HEALTH ORGANISATION. 2006. *Body mass index (BMI) classification* [Online]. World Health Organisation. [Accessed].
- WRIGHT, A. A., COOK, C. & ABBOTT, J. H. 2009. Variables associated with the progression of hip osteoarthritis: a systematic review. *Arthritis & Rheumatism*, 61, 925-36.
- WRIGHT, A. A., COOK, C. E., FLYNN, T. W., BAXTER, G. D. & ABBOTT, J. H. 2011. Predictors of response to physical therapy intervention in patients with primary hip osteoarthritis. *Physical Therapy*, 91, 510-24.
- YAMADA, H., KOSHINO, T., SAKAI, N. & SAITO, T. 2001. Hip adductor muscle strength in patients with varus deformed knee. *Clinical Orthopaedics & Related Research*, 179-85.
- ZHANG, W., DOHERTY, M., PEAT, G., BIERMA-ZEINSTRAS, M. A., ARDEN, N. K., BRESNIHAN, B., HERRERO-BEAUMONT, G., KIRSCHNER, S., LEEB, B. F., LOHMANDER, L. S., MAZI, X00E, RES, B., PAVELKA, K., PUNZI, L., SO, A. K., TUNCER, T., WATT, I. & BIJLSMA, J. W. 2010. EULAR evidence-based recommendations for the diagnosis of knee osteoarthritis. *Annals of the Rheumatic Diseases*, 69, 483-9.
- ZHANG, W., MOSKOWITZ, R. W., NUKI, G., ABRAMSON, S., ALTMAN, R. D., ARDEN, N., BIERMA-ZEINSTRAS, S., BRANDT, K. D., CROFT, P., DOHERTY, M., DOUGADOS, M., HOCHBERG, M., HUNTER, D. J., KWOH, K., LOHMANDER, L. S. & TUGWELL, P. 2008. OARSI recommendations for the management of hip and knee osteoarthritis, Part II: OARSI evidence-based, expert consensus guidelines. *Osteoarthritis & Cartilage*, 16, 137-62.

Protocol for Conducting Systematic Review

Title

“Patient characteristics that predict long term outcome of knee osteoarthritis: A systematic review of prognostic studies.”

Methods

Search strategy

- Electronic Databases: MEDLINE, EMBASE, CINAHL, AMED, Cochrane Central registry (CENTRAL)
- Citation database: Web of Science and Web of Knowledge
- Hand search: reference lists of selected (final) papers and relevant reviews

Record years searched and number of articles retrieved, date of search, person searching.

Search Terms

For Ovid MEDLINE, (will be adapted to other databases)

1. Knee joint; osteoarthritis; osteoarthritis, knee; knee arthrosis; arthralgia
2. Best Sensitivity = incidence.sh. OR exp mortality OR follow-up studies.sh. OR prognos:.tw. OR predict:.tw. OR course:.tw.
3. Best optimization of sensitivity & specificity = prognosis.sh. OR diagnosed.tw. OR cohort:.mp. OR predictor:.tw. OR death.tw. OR exp models, statistical {Wilczynski, 2004 #2724}

The search strategy will be piloted to ensure good coverage of key articles.

Two reviewers will screen the titles and abstracts according to inclusion/exclusion criteria. Reasons for exclusions will be documented. If there is disagreement between the two reviewers, they will meet to discuss and

try to reach agreement. A third reviewer will be available to resolve differences.

Inclusion/Exclusion criteria

Inclusion

Participants

- Human adults with OA knee diagnosis
- Any age (>18), any gender, any duration of symptoms

Interventions

- Any exposure of cohort was identified
- If RCT must be evidence of no effect of intervention or placebo group reported separately

Comparison

- Comparison allowed OA knee with no OA knee, or with treatment arm of RCT if evidence of no effect.

Outcome

- Progression of OA knee defined as:-
- Change in functional status/pain reporting (WOMAC, VAS, NPRS etc)
TKR
- Xray change (e.g. change in grade Kellegren-Lawrence, joint space narrowing)
- (MRI change/Biomarker change as long as in addition to one of the above)

Study design

- The study must identify prognostic factors at baseline and report a statistical association or lack of association with outcome
- Self reported pain and/or function and/or patient characteristics (e.g.age/BMI etc) must be reported at baseline
- Longitudinal cohort study

- RCT if placebo group with OA knee reported separately
- Follow-up \geq 1 year

Exclusion

- Underlying pathology such as RA/Ca/osteoporosis/joint infection
- Animal studies
- Generalised OA where knee OA results not reported separately
- Previous surgery for OA knee (TKR, osteotomy, arthroscopic lavage)
- Duplicate report
- Any study that focuses entirely on xray changes/ MRI changes or changes in biomarkers without reference to clinical presentation/features such as pain/stiffness/loss of function or patient characteristics such as age/BMI.

Criteria will be piloted on ten key articles. Full text articles will be obtained.

The same two raters will apply inclusion/exclusion criteria. Reasons for exclusion will be documented. Disagreement will be resolved as above.

Data extraction and analysis

A data extraction form will be developed to extract data from full text articles.

This will be performed by two reviewers.

Assessment of Study quality

Use criteria based approach rather than score, as scores can be arbitrary, unreliable and hard to interpret {Altman, 2001 #2736}; i.e. criteria present or absent (positive/negative/unclear) {van Tulder, 2003 #5432}. Develop modified checklist based on previous similar studies. Pilot checklist with two reviewers, on five studies. Document reasons for any exclusions.

Agreement will be reported using kappa statistic.

Synthesis of evidence

It is unlikely that in a review of prognostic studies there will be sufficient homogeneity to pool results and perform a meta-analysis {Altman, 2001 #2736}. Some studies will report OR/RR/HR and it may be possible to calculate these from other studies if data is sufficient. An alternative method for grading of evidence is to be decided, although it is likely to be a narrative synthesis. Sensitivity analyses will be performed to determine whether increasing quality score defining high quality studies, alters the overall findings of the review

Search terms for Medline (modified for other databases)

1. incidence.sh.
2. exp mortality/
3. follow-up studies.sh.
4. prognos:.tw.
5. predict:.tw.
6. course:.tw.
7. 6 or 4 or 1 or 3 or 2 or 5
8. exp osteoarthritis/
9. osteoarthritis.tw.
10. exp arthritis/
11. arthritis.tw.
12. arthrosis.tw.
13. 8 or 11 or 10 or 9 or 12
14. exp knee/
15. 13 and 14
16. exp osteoarthritis, knee/
17. osteoarthritis, knee.tw.
18. gonarthrosis.tw.
19. 18 or 16 or 17 or 15
20. 7 and 19
21. 20 not surgery.mp.[mp=title, original title, abstract, name of substance word, subject heading word]
22. 21 not arthroplasty.mp. .[mp=title, original title, abstract, name of substance word, subject heading word]
23. 22 not osteotomy.mp. [mp=title, original title, abstract, name of substance word, subject heading word]
24. limit 23 to "all adult (19 plus years)"

Risk of Bias Assessment for Prognostic Studies of OA Knee

Criteria	Methodological Quality	Yes	No	Unsure
Study Participation				
A	Inception cohort. Were all subjects recruited at an early uniform point in their disease? Duration of symptoms/stage of disease described? Positive if similar for whole cohort or if cohort grouped according to similar stage of disease.			
B	Description of study population. Was the setting of recruitment described? (e.g primary care/specialist clinics) Are baseline characteristics of study sample adequately described? (including possible prognostic factors) Positive if can determine that cohort is representative of source population.			
C	Description of inclusion and exclusion criteria . Did Inclusion/Exclusion criteria address age, duration of symptoms, diagnostic criteria, relevant comorbidities?			
D	Response of $\geq 75\%$ for cohorts and controls. At least 75% of people eligible participated in trial.			
Study attrition				
E	Follow-up of at least 12 months			
F	Dropouts/loss to follow-up <20%			
G	Information completers vs. loss to follow-up/dropouts. (reasons for withdrawal, any differences in key characteristics or outcomes between dropouts and completers)			
Measurement and data presentation				
H	Prospective data collection			
I	Treatment/Exposure in cohort is fully described/standardized.			
J	Clinically relevant prognostic factors.			
K	Standardized or valid measures. Are the prognostic factors clearly			

Appendix C: Operational definitions for assessment of bias tool

	defined and reliably measured? (limited reliance on recall, measurement method described, setting and method of measurement same for all subjects)			
L	Data presentation of the most important prognostic factors.			
M	Clinically relevant outcome measures. Did the outcome measures assess pain, function, stiffness or radiologic changes? Was MCID defined?			
N	Standardized or valid measurements. Use of recognised criteria? Was measurement of outcome standardised and reliable with limitation of bias (standardised methods, blinded assessors, method/setting same for all subjects) Was progression defined? Did the outcome measure selected answer the question “was there progression of OA in the knee?”			
O	Data presentation of the most important outcome measures. Is there adequate reporting of results? (avoiding selective reporting, sufficient presentation of data to assess the adequacy of the analysis, frequency/percentage/mean/median/SD/ CI reported for outcome measures and prognostic factors)			
Analysis and presentation of results				
P	Appropriate analysis techniques. Is the statistical analysis appropriate for the design of the study? (appropriate model selected e.g. multiple regression, calculation of OR/RR/HR, measures of association/variance)			
Q	Prognostic model is presented. Numerical description of important outcomes given.			
R	Sufficient Numbers. Sample size calculation? Is it adequate in relation to number of prognostic factors 10:1? >100 = positive.			
S	Confounding Were possible confounders identified and accounted for in the analysis?			

Appendix C: Operational definitions for assessment of bias tool

	(matching for key variables, stratification, initial assembly of comparable groups, exposure to interventions) e.g. age/sex/BMI			
T	Conclusions Discussion of limitations of study. Were conclusions supported by results?			
MCID = minimally important clinical difference; OA = osteoarthritis; OR = odds ratio; RR = relative risk; HR = hazard ratio; BMI = body mass index				

Appendix D: Data extraction form for studies included in systematic review

Methods

Outcome measure to establish progression:

Definition of OA progression:

Description of exposure (if any):

Statistical Analysis

Statistics used to report associations of prognostic factor & outcome:

Results

Prognostic factors identified/studied:

Summary data of main findings (OR/RR/HR/correlation) including statistical significance:

State covariates/confounders:

Adjustment for covariates/confounders:

Main Conclusion

State main conclusion(s):

Record funding type and source:

Table A Systematic review of predictors of knee osteoarthritis progression: study details.

Reference	Study Population	Follow-up in months	Number of participants	% female	Age (years) (mean±SD)	OA diagnosis for inclusion	Outcome measure	Prognostic factors investigated	Quality score (/20)
Amin et al (2009)(20)	Boston Osteoarthritis of the Knee study (BOK). Veterans Administration clinics, community.	15 and 30	265 (@15/12) 222 (@30/12)	42	67 ± 9	ACR Osteophyte on xray	Change in: Pain (VAS), Function (WOMAC)	Quadriceps strength	15
Amin et al (2008)(21)	BOK study. Veterans Administration clinics, community.	15 and 30	265 (@15/12) 223 (@30/12)	43	67 ± 9	ACR Osteophyte on xray	Change in: Pain (VAS), Function (WOMAC)	ACL tear	16
Amin et al (2007)(22)	BOK study. Veterans Administration clinics, community.	15 and 30	159 (@15/12) 133 (@30/12)	nil	68 ± 9	ACR Osteophyte on xray	Change in: Pain (VAS)	Cigarette smoking	14
Benichou et al (2010) (23)	Osteoarthritis Initiative (OAI) study. Community based cohort	12	67	64	60 ± 9	JSN (OARSI grade 1-3)	Change in JSW	BMI > 25kg/m ² , age, sex, osteophytes, baseline JSW, history of knee injury	18
Bruyere et al (2003)(24)	Subjects from glucosamine sulphate RCT	36	71 in placebo group	Not given	Not given	ACR	JSN>0.5mm	JSW @ baseline, age, BMI, WOMAC	10

Appendix E: Details of studies included in systematic review

Reference	Study Population	Follow-up in months	Number of participants	% female	Age (years) (mean±SD)	OA diagnosis for inclusion	Outcome measure	Prognostic factors investigated	Quality score (/20)
Cooper et al (2000)(25)	Registered @ GP practice, community	60	354	72	75.8 (median) 69.5-80.9 (IQ range)	KL grade, knee pain	↑KL grade	Knee pain, Heberden's nodes, previous knee injury, sports participation, obesity	15
Dieppe et al (1997)(26)	OA500 study, rheumatology clinic patients	36	a)145 for clinical variables b)138 for radiographic variables	67.5	62.6 ±11.7	Radiographic evidence	Joint space change ≥2mm (tibiofemoral), Change in grade sclerosis or osteophytes	a) Age, sex, BMI, CRP, RhF, Rest pain, pain at entry, Steinbroker at entry, b) Entry joint space, radiological features in different compartments	13
Dieppe et al (1993)(27)	Rheumatology clinic	60	60 14 progressed with JSN ≥2mm (other 22 had joint surgery)	69	64.2 ±11.6	Knee pain, radiographic evidence (JSN, osteophyte)	tibiofemoral JSN ≥2mm	Bone scan abnormalities, age, sex, BMI, knee pain, crepitus, bony swelling, soft tissue swelling, instability, walking ability, duration of symptoms, radiographic variables	10
Dougados et al (1992)(28)	Rheumatology patients	12	360	71	67 ±10	ACR	Tibiofemoral JSN assessed by 1.change in grade (6 grade scale) 2. distance in mm	Age, sex, obesity, intermittent pain, chondrocalcinosis, varus/valgus, medial joint space involvement, generalised OA, no of OA joints	10

Appendix E: Details of studies included in systematic review

Reference	Study Population	Follow-up in months	Number of participants	% female	Age (years) (mean±SD)	OA diagnosis for inclusion	Outcome measure	Prognostic factors investigated	Quality score (/20)
Felson et al (2004)(29)	BOK study, Veterans Administration clinics, community	15 and 30 months	227 90 knees showed progression	41	66.4 ± 9.4	ACR	Tibiofemoral JSN Change in grade (0-3)	BMI	16
Golightly et al (2010) (30)	Johnston County OA project Community based, 6 townships, USA	69(median)	N = 1583 1657 knees with OA @ baseline 614 progressed	64.3	60.9 ± 10	KL grade ≥ 1	↑ KL grade	Leg length inequality (LLI)	17
Harvey et al (2010) (31)	Multicentre Osteoarthritis Study (MOST), 2 communities, USA	30	2964	LLI <1cm 61.6 LLI ≥ 1cm 53.4	LLI <1cm 62.4 ± 8.1 LLI ≥ 1cm 63.1 ± 8.0	KL grade ≥ 2	↑ JSN score TKR	Leg length inequality (LLI)	17
Hellio Le Graverand et al (2008) (32)	MOST study, obese females	12	60	100	KL grade 2 55.5 ± 7.4 KL grade 3 58.2 ± 8.3	KL grade ≥ 2	JSN	BMI, baseline KL grade	16

Appendix E: Details of studies included in systematic review

Reference	Study Population	Follow-up in months	Number of participants	% female	Age (years) (mean±SD)	OA diagnosis for inclusion	Outcome measure	Prognostic factors investigated	Quality score (/20)
Ledingham et al (1995)(33)	Rheumatology clinic	24	188 252 knees showed at least one sign of progression	62.8	70 range (34-91)	JSN and osteophyte	↑ KL grade ↑ any radiological feature Worsening of symptoms (global) ↑ pain ↑ disability (Steinbrocker). ↓ exercise tolerance	Nodal change, polyarticular IP OA, multiple joint OA, synovial fluid volume, effusion, warmth, CPPD, severity of OA, age, BMI, chondrocalcinosis, pain score	15
Mazucca et al (2006)(34)	Derived from 2 cohorts (one female placebo group from RCT), variety of community and clinical settings	30	288 107 knees progressed by JSN 129 knees progressed by osteophytosis	84	60 ±9.6	Radiographic evidence of marginal tibiofemoral osteophyte + JSW ≥ 2mm. RCT cohort – KL criteria	↑ osteophytosis ↑ JSN medial or lateral joint space	Age, sex, race, BMI, duration of symptoms, duration of diagnosis, WOMAC pain, WOMAC stiffness, WOMAC function, JSW medial T/F jt, osteophyte, Presence PF OA, contralateral OA	16
Mazucca et al (2005)(35)	Placebo group of RCT, obese women with unilateral OA knee	16 and 30	73 @ 16 months 70 @ 30 months Progressors 17 @ 16 months 23 @ 30 months	100	55.6 ±5.7	Index knee -KL grade 2/3 Contralateral knee – KL grade 0/1	Medial tibiofemoral joint JSN ≥ 0.5mm	Bone scintigraphy, baseline JSW, radiographic features, knee pain	14

Appendix E: Details of studies included in systematic review

Reference	Study Population	Follow-up in months	Number of participants	% female	Age (years) (mean±SD)	OA diagnosis for inclusion	Outcome measure	Prognostic factors investigated	Quality score (/20)
McAlindon et al (1996) ^a (36)	Framingham study, population based cohort Examinations 18 and 22	> 84	126 62 knees progressed	N/S	70.3 ±4.5	KL grade ≥ 2	↑ KL grade, osteophyte growth, JSN (cartilage loss)	Dietary intake of Vitamin D, serum levels of Vitamin D	16
McAlindon et al (1996) ^b (37)	Framingham study, population based cohort Examinations 18 and 22	> 84	187 68 knees progressed	64.2	70.3 ±4.5	KL grade ≥ 2	↑ KL grade	Antioxidant nutrients – vitamin C, beta carotene, vitamin E. Nonantioxidant nutrients – vitamins B ₁ , B ₆ , niacin and folate	16
Niu et al (2009)(38)	MOST study, community, multicentre	30	2623 54% progressed	59.4	62.4 ±8.0	Individuals with OA – KL grade 2 or 3 High risk individuals- obesity, knee pain, history of knee injury or previous knee surgery	↑ JSN ≥ one-half grade in any location	BMI (obesity)	16
Reijman et al (2007)(39)	Rotterdam study, population cohort	72	532 Progressors 21.8%, JSN ≥ 1mm 8.1% JSN ≥ 1.5mm 11.4% ↑ KL grade	68.4	68.6 ±7.0	KL grade	JSN ≥ 1mm JSN ≥ 1.5mm ↑ KL grade	BMI	15

Appendix E: Details of studies included in systematic review

Reference	Study Population	Follow-up in months	Number of participants	% female	Age (years) (mean±SD)	OA diagnosis for inclusion	Outcome measure	Prognostic factors investigated	Quality score (/20)
Schouten et al (1992)(40)	Population cohort	144	142 48 progressed	59	57.2 ±6.1	KL grade ≥2	↓ JSW	Age, BMI, weight, uric acid concentration, gender, jogging or member of sports club, chondrocalcinosis, Heberden's nodes, generalised OA, localised OA, smoking (have not included variables assessed only at follow-up)	15
Sharma et al (2010) (41)	MOST study 2 communities, USA	30	N = 950 1307 knees analysed	62	63.6 ± 7.8	KL grade ≥2	↑ JSN grade	Varus/valgus alignment	19

Appendix E: Details of studies included in systematic review

Reference	Study Population	Follow-up in months	Number of participants	% female	Age (years) (mean±SD)	OA diagnosis for inclusion	Outcome measure	Prognostic factors investigated	Quality score (/20)
Sharma et al (2003) ^a (42)	Mechanical factors in arthritis of the knee (MAK) study, longitudinal cohort, community, multiple sources	36	236 123 poor outcome 156 poor outcome chair-stand performance	73	68.6 ±10.8	KL ≥ 2, WOMAC (function) NIAMS/NIA	Move to lower function group or remain in 3 lowest-functioning groups for WOMAC (function) and chair-stand performance	BMI, knee pain intensity (VAS), varus-valgus laxity, alignment, mental health score, self-efficacy score, social support score, aerobic exercise score, age, proprioceptive inaccuracy, quadriceps muscle strength, hamstrings muscle strength, role functioning emotional score Sex, marital status, comorbidity, radiographic severity OA, bilateral v unilateral OA	18
Sharma et al (2003) ^b (43)	MAK study, longitudinal cohort, community, multiple sources	18	171 Used 328 knees for analysis	73.6	64 ±11.0	KL ≥ 2, WOMAC (function) NIAMS/NIA	↑ grade JSN any compartment	Quadriceps strength, alignment	18

Appendix E: Details of studies included in systematic review

Reference	Study Population	Follow-up in months	Number of participants	% female	Age (years) (mean±SD)	OA diagnosis for inclusion	Outcome measure	Prognostic factors investigated	Quality score (/20)
Sharma et al (2001)(44)	MAK study, longitudinal cohort, community, multiple sources	18	230 37 knees progressed medially 24 progressed laterally	75.2	64 ±11.1	KL ≥ 2, WOMAC (function) NIAMS/NIA	JSN for medial progression JSN for lateral progression ↑ KL grade ↓ chair-stand rate ↓ 20% in rate of chair-stand performance (function)	Varus alignment, valgus alignment,	18
Shiozaki et al (1999)(45)	Adult women from the Matsudai Knee Osteoarthritis Survey, population cohort	168	608 432 knees progressed	100	51.3 Range 40 - 65	KL grade 1 or 2	↑ KL grade	BMI	13
Spector et al (1994)(46)	Chingford population cohort. Sub-sample of women with unilateral knee OA	24	118 @ baseline 58 used in analysis	100	56.8 ±5.9	KL ≥ 2	↑ KL grade ↑ osteophyte size JSN	BMI, age, physical activity, trauma, hand OA, knee pain, weight	16
Spector et al (1992)(47)	2 cohorts of patients with OA hand and knee: rheumatology clinic patients, subjects from previous clinical trial	132	63 13 progressed ≥ 1 KL grade	76.2	69 Range 52-87 (@ F/U)	KL grade ≥ 1	↑ KL grade JSN >10% TKR	OA severity, knee pain, sex, weight change	8

Appendix E: Details of studies included in systematic review

Reference	Study Population	Follow-up in months	Number of participants	% female	Age (years) (mean±SD)	OA diagnosis for inclusion	Outcome measure	Prognostic factors investigated	Quality score (/20)
Thorstensson et al (2004)(48)	Population cohort, urban and rural community	60	54 29 progressed	41.9	44.8 Range 35-54	KL ≥ 1	↑ KL grade	Age, sex, BMI, knee pain, 3 tests of lower extremity functional performance (max no of 1-leg rises from sitting, 300m timed walk, timed 1-leg stand)	15
Wolfe & Lane (2002)(49)	OA hip or knee patients at arthritis clinic	42 - 81	1232	77	63.4 ±11.8	ACR clinical criteria	Progression to max JSN (= 3)	Initial JSN score, global severity, BMI, pain, symptom duration, ESR, WBC, HAQ disability, sex, depression, level of education, anxiety, haemoglobin, age	15
<p>OA = osteoarthritis, SD = standard deviation, ACR = American College of Rheumatology classification criteria , VAS = visual analogue scale, WOMAC = Western Ontario and Macmaster University Osteoarthritis index, , ACL = anterior cruciate ligament, JSN = joint space narrowing, JSW = joint space width, RCT = randomized control trial, BMI = body mass index, GP = general practitioner, KL = Kellgren Lawrence grade, IQ range = inter quartile range, CRP = C reactive protein, RhF = rheumatoid factor, TKR = total knee replacement, IP OA = interphalangeal osteoarthritis, SF = synovial fluid, CPPD = calcium pyrophosphate crystals, T/F jt = tibio femoral joint, PF OA = patello femoral osteoarthritis, N/S = not stated, F/U = follow-up, NIAMS/NIA = National Institute of Arthritis and Musculoskeletal and Skin Diseases/National Institute on Aging, ESR = erythrocyte sedimentation rate, WBC = white blood count, HAQ disability = Stanford Health Assessment Questionnaire functional disability index.</p>									

Table B Demographic and patient-reported variables: association with progression of knee osteoarthritis

Variable	Ref no†	Outcome Measure	Study Results – significant effect on progression	Study Results- no significant effect on progression
<i>Demographic information</i>				
Age	40	JSW	> 60 years OR = 3.84 (1.10 – 13.4)**	50 – 54 years OR = 2.21 (0.57 – 8.66)** 55 – 59 years OR = 1.94 (0.49 – 7.61)** OR 1.57 (0.53 – 4.60)*
	48‡	↑KL grade		
	42	Chair stand performance	OR = 1.34 (1.15-1.57)***	
	34	JSN		OR 1.13 (0.87 – 1.48)**
	33	↑disability∞	OR 1.07 (1.02 – 1.13)***	
	23	JSW	$\beta = 0.01 \pm 0.006\text{mm/year}$ $p = 0.03$ ***	
	26‡ (24) (28)	Self-report worsening§ JSW JSN	↑ age and ↓ function more likely to progress *** ††	$p = 0.36$ *** $p = 0.302$ *
Sex (Female : male unless otherwise stated)	33	↑cyst	OR 2.17 (1.13 – 4.15)***	
	40	JSW		OR 0.5 (0.22 – 1.11)**
	48‡	↑KL grade		OR 1.51 (0.51 – 4.47)*
	26‡	Self-report worsening§	male/↓pain more likely to improve than female/↑pain***	female/↑rest pain/↓joint pain more likely to progress*** No significant difference in outcome*
	42	WOMAC (F) Chair stand performance		$p = 0.763$ *
	(28) (47)	JSN ↑KL grade		$p = 0.3$ *
Psychosocial factors	42	WOMAC (F)	Mental health score OR 0.58 (0.39 – 0.86)*** Self-efficacy	

Appendix F: Table B and Table C, results extracted from all studies in systematic review

Variable	Ref no†	Outcome Measure	Study Results – significant effect on progression	Study Results- no significant effect on progression
		Chair stand performance	OR 0.80 (0.65-0.98)*** Social support OR 0.85 (0.73 – 0.98)***	Role-functioning OR 0.99 (0.75 – 1.32)*** Self-efficacy OR 0.86 (0.68 – 1.09)***
Marital Status	42	WOMAC (F) Chair stand performance		No difference in outcome*
<i>Symptoms</i>				
Pain	42 26‡ 25 48‡ (24) (28) (47)	WOMAC (F) Self-report worsening§ ↑KL grade ↑KL grade JSW JSN JSN	male/↓pain more likely to improve than female/↑pain***	OR 1.12 (0.90 – 1.40)*** female/↑rest pain/↓joint pain more likely to progress*** KL grade 1 OR 0.8 (0.4 – 1.7)** KL grade 2 OR 2.4 (0.7 – 8.0)** OR 2.33 (0.66 – 8.30)* p = 0.81*** p = 0.06* p = 0.2*
Function	26‡ 34 48‡	Self-report worsening§ JSN ↑KL grade	↑ age and ↓ function more likely to progress ***	OR 1.16 (0.92 – 1.47)** Max 1-leg raise OR 1.33 (0.46 – 3.90)* 300m walk OR 0.98 (0.32 – 3.02)* 1-leg stand

Appendix F: Table B and Table C, results extracted from all studies in systematic review

Variable	Ref no†	Outcome Measure	Study Results – significant effect on progression	Study Results- no significant effect on progression
	(24)	JSN		OR 1.56 (0.53 – 4.60)* p = 0.51***
Joint Stiffness	34 (24)	↑Osteophytes JSN	OR 1.39 (1.09-1.77)**	p = 0.29***
Severity of disease (clinical)	49	JSN	HR 1.01 (1.00 – 1.02)***	
Disease history				
Previous knee injury	25	↑KL grade		KL grade 1 OR 1.2 (0.5 – 3.0)** KL grade 2 OR 1.1 (0.3 – 4.4)**
	21	Change in pain WOMAC (F)		p > 0.05** p > 0.05**
Symptom Duration	49	JSN	HR 1.3 (1.00 – 1.05)***	
Patient behaviour/actions				
Activity levels/sports participation	42	WOMAC (F) Chair stand performance		OR 0.84 (0.69 – 1.02)*** OR 0.86 (0.71 – 1.05)***
	25	↑KL grade		KL grade 1 OR 0.7 (0.4 – 1.6)** KL grade 2 OR 0.9 (0.3 – 2.5)**
	40	JSW		OR 0.53 (0.17 – 1.68)**
Smoking	40	↓ JSW		Previous smoker OR = 1.07 (0.38 – 3.04)** Current smoker OR = 0.96 (0.34 – 2.75)**
	22	Change in pain (VAS)		p > 0.05**
↓Vitamin D intake	36	↑KL grade	OR 2.99 (1.06 – 8.49) to	

Appendix F: Table B and Table C, results extracted from all studies in systematic review

Variable	Ref no†	Outcome Measure	Study Results – significant effect on progression	Study Results- no significant effect on progression
Antioxidant intake (stratified by sex)	37	↑KL grade	<p>OR 4.05 (1.4 – 11.6)***</p> <p>Vitamin C/upper tertile Male OR 0.06 (0.01 – 0.67)*** Female OR 0.31 (0.12-0.79)*** Vitamin C/middle tertile Female OR 0.36 (0.14 – 0.90)***</p> <p>Vitamin E/upper tertile Male OR 0.07 (0.01 – 0.61)***</p>	<p>Vitamin C/middle tertile Male OR 0.13 (0.00-3.81)</p> <p>Beta carotene/all groups OR 0.26 (0.04 – 1.57) to OR 1.75 (0.74 – 4.11)*** Vitamin E/upper tertile Female OR 0.48 (0.40-2.90) Vitamin E/middle tertile Male OR 0.01 (0.00 – 1.37)*** Female OR 0.81 (0.32 – 2.03)***</p>
<p>Results are ordered by number of studies investigating characteristic first; then by strength of findings and quality of studies. †high quality studies in bold, low quality studies in parentheses; OR = odds ratio; HR = hazard ratio; *Univariate analysis – not adjusted; **univariate analysis - adjusted; *** multivariate analysis; ∞ = Steinbrocker functional index; ‡ = data not given for adjusted results or multivariate models; § = outcome measure one of: ↑pain/↓functional status/walking aid/major joint surgery; JSW = joint space width; JSN = joint space narrowing; ↑KL grade = increase in Kellgren/Lawrence grade, 1+ by one grade, 2+ by two grades; WOMAC = Western Ontario and McMaster Universities Osteoarthritis Index, (F) = function subscale; VAS = visual analogue scale</p>				

Table C Physical examination and clinical tests: association with knee osteoarthritis progression

Variable	Ref not†	Outcome Measure	Study Results – significant effect on progression	Study Results- no significant effect on progression
<i>Physical Examination</i>				
BMI / Obesity	40	JSW	25.97-27.73 kg/m ² OR 5.28 (1.54-18.1)** > 27.3 kg/m ² OR 11.1 (3.28-37.3)**	24.35-25.96 kg/m ² OR 1.77 (0.48-6.5)**
	46	↑osteophytes		RR 4.69 (0.63 – 34.75)**
	39	JSN ≥ 1mm		> 25 – 27.5 kg/m ² OR 1.2 (0.6-2.4)** > 27.5 kg/m ² OR 1.4 (0.8-2.6)**
		JSN ≥ 1.5mm	> 27.5 kg/m ² OR 3.2 (1.1-9.7)**	> 25 – 27.5 kg/m ² OR 2.3 (0.7-7.7)**
		↑ KL grade	> 27.5 kg/m ² OR 2.1 (1.2-3.7)**	> 25 – 27.5 kg/m ² OR 1 (0.5-2.0)**
	25	↑KL grade	KL grade 1/>25.4 kg/m ² OR 2.6 (1.0-6.8)**	KL grade 1/22.7-25.4 kg/m ² OR 2.3 (0.8-6.4)** KL grade 2/22.7-25.4 kg/m ² OR 1.8 (0.4-8.2)** KL grade 2/>25.4 kg/m ² OR 1.3 (0.3-5.0)**
	48	↑KL grade		OR 1.87 (0.62-5.63)*
	45‡		25.0-26.6 kg/m ² RR 1.38 (1.10-1.73)* 26.7-34.7 kg/m ² RR 1.51 (1.22-1.87)*	17.3-21.4 kg/m ² RR 0.81 (0.60-1.10)* 21.5-23.2 kg/m ² RR 1.00* 23.3-24.9 kg/m ² RR 1.05 (0.8-1.36)*
	42	WOMAC (F)		OR 1.14 (0.89-1.46)***
	33			OR 1.07 (1.02-1.14)***

Appendix F: Table B and Table C, results extracted from all studies in systematic review

Variable	Ref no†	Outcome Measure	Study Results – significant effect on progression	Study Results- no significant effect on progression
BMI- stratified by alignment	49	JSN	OR 1.06 (1.00-1.12)***	
		Osteophytes	HR 1.03 (1.00 – 1.06)***	
	23	JSN	$\beta = -0.03$	
		JSW	($\pm 0.01\text{mm}$) per kg/m^2 $p = 0.03$	
	32			$\geq 30 \text{ kg}/\text{m}^2$
	(24)	JSN		$p = 0.256^*$
	(27)	JSN	$p < 0.01^*$	$p = 0.18^{***}$
	(28)	Interbone distance	$p = 0.021^*$	
	29	JSN	Female/Moderate malalignment	Female/Neutral alignment
		JSN	OR 1.39 (1.07-1.80)**	OR 1.00 (0.83-1.20)**
38	JSN	Neutral alignment/ BMI $> 35 \text{ kg}/\text{m}^2$ RR 1.8 (1.0-3.2)***	Female/Severe malalignment	
			OR 0.96 (0.68-1.36)**	
			Male/Neutral alignment	
			OR 0.96 (0.72-1.26)**	
			Male/Moderate malalignment	
			OR 1.14 (0.92-1.40)**	
			Male/Severe malalignment	
			OR 0.92 (0.71-1.18)**	
			Neutral alignment/ BMI 25-29.9 kg/m^2	
			RR 1.2 (0.7-2.1)***	
			Neutral alignment/ BMI 30-34.9 kg/m^2	
			RR 1.2 (0.7-2.2)***	
			All alignment/ all BMI	
			RR 1.0 (0.8-1.2) to	
			RR 1.0 (0.9-1.4)***	
			Varus alignment/ all BMI	

Appendix F: Table B and Table C, results extracted from all studies in systematic review

Variable	Ref no†	Outcome Measure	Study Results – significant effect on progression	Study Results- no significant effect on progression
				RR 0.9 (0.7-1.1)*** Valgus alignment/ all BMI RR 0.8 (0.6-1.4) to RR 1.4 (0.9-2.1)***
Heberden's nodes/ Bony swelling/ Bouchard's nodes	40 33 25	JSW ↑ KL grade ↑ KL grade	OR 5.97 (1.54-23.1)** OR 1.80 (1.02-3.17)***	KL grade 1 OR 0.7 (0.4-1.6)** KL grade 2 OR 2.0 (0.7-5.7)**
Muscle strength	42 43 20	↓ Chair stand performance JSN (compared PP high to low strength tertiles) Change in pain (VAS) Change in WOMAC (F)	Malaligned (≥5°) p = 0.03**	Quadriceps OR 0.88 (0.70-1.11)*** Hamstrings OR 0.86 (0.60-1.23)*** Neutral alignment (<5°) p > 0.2** Malaligned (≥5°)/ BMI<30kg/m ² p = 0.06** p = 0.4** p = 0.5**
Leg length inequality	30 31	↑ KL grade ↑symptoms JSN score	KL grade 2 HR 1.83 (1.10-3.05)*** Short leg/≥1cm OR 1.3 (1.0-1.7)***	KL grade1 HR 1.22 (0.82-1.80)*** HR 1.13 (0.53-2.39)*** Short leg/≥2cm OR 1.4 (0.5-3.7)*** Long leg/≥1cm OR 1.2 (0.91-1.5)*** Long leg/≥2cm OR 1.0 (0.4-2.5)***
Warmth	33	↑ symptoms ↑ disability	OR 2.14 (1.30 – 3.52)*** OR 4.25 (1.66 – 10.9)***	

Appendix F: Table B and Table C, results extracted from all studies in systematic review

Variable	Ref no†	Outcome Measure	Study Results – significant effect on progression	Study Results- no significant effect on progression
		↓ exercise tolerance Radiographic features	OR 2.24 (1.32 – 1.80)*** OR 2.22 (1.19 – 4.14)*** to OR 3.03 (1.63 – 5.64)***	
↑Volume synovial fluid (aspirated)	33	Radiographic features	OR 1.02 (1.00-1.05) to OR 1.03 (1.01-1.05)***	
Crepitus	(27)	JSN ≥ 2mm	p < 0.01*	
Instability	(27)	JSN ≥ 2mm	p < 0.001*	
<i>Radiographic examination</i>				
Alignment	44	Medial JSN	Varus alignment OR 2.98 (1.51-5.89) to OR 4.09 (2.20-7.62)**	
		Lateral JSN	Valgus alignment OR 3.42 (1.31-8.96) to OR 4.89 (2.13-11.20)**	
		↑ KL grade	Varus alignment OR 3.61 (1.33-9.85)**	Valgus alignment OR 2.51 (0.91-6.89)**
		↓ chair-stand rate		Unilateral malalignment OR 0.17 (-1.66 to 2.01)** Bilateral malalignment OR 2.23 (0.05 to 4.41)**
		↓ chair-stand rate by 20%	Bilateral malalignment OR 3.22 (1.28-8.12)**	Unilateral malalignment OR 2.33 (0.97-5.62)**
	41	Medial JSN	Varus OR 3.59 (2.62, 4.92)*** OR 4.21 (2.84, 6.24) female OR 2.75 (1.63, 4.66) male	
			Valgus OR 0.34 (0.21, 0.55)*** OR 0.37 (0.21, 0.65) female	OR 0.34 (0.11, 1.03) male
		Lateral JSN	Varus OR 0.12 (0.07, 0.21)*** OR 0.15 (0.06, 0.33) female	

Appendix F: Table B and Table C, results extracted from all studies in systematic review

Variable	Ref no†	Outcome Measure	Study Results – significant effect on progression	Study Results- no significant effect on progression
	42 (28)	WOMAC (F) Chair stand performance JSN	OR 0.08 (0.04, 0.19) male Valgus OR 4.85 (3.17, 7.42)*** OR 6.35 (3.76, 10.75) female	OR 1.88 (0.69, 5.12) male No significant difference* p = 0.923*
Baseline JSW/ JSN/ joint space/ interbone distance	34 49 (24) (27)	JSN JSN JSN TKR	OR 0.63 (0.47-0.86)*** HR 2.53 (1.94 – 3.31)*** p < 0.001*	RR 2.39 (0.99-5.79)***
Severity of OA (KL grade)	33 32 42	Radiographic features ↑ global symptoms ↑ pain ↓ exercise tolerance JSN WOMAC (F) Chair stand performance	OR 1.72 (1.36 – 2.19)*** OR 1.29 (1.07 – 1.57)*** OR 1.29 (1.08 – 1.55)*** OR 1.52 (1.22 – 1.88)*** KL grade2 p < 0.005* KL grade 3 p < 0.001*	No significant difference*
Chondro-calcinosis	40 (28)	JSW JSN		OR 2.01 (0.55 – 7.42)** p = 0.832*
High baseline osteophyte score	34	Osteophyte growth	OR 0.47 (0.33 – 0.66)**	
<i>Pattern of joint involvement</i>				
Ipsilateral PF OA	34	JSN Osteophyte growth	OR 3.36 (1.83-6.18)*** OR 2.31 (1.37 – 3.88)**	
Contra-lateral knee OA	34	JSN		OR 1.53 (0.82 – 2.85)**
Localised OA	40	JSW		OR 1.17 (0.51 – 2.72)**
Bilateral knee OA	42	WOMAC (F)		No significant difference*

Appendix F: Table B and Table C, results extracted from all studies in systematic review

Variable	Ref no†	Outcome Measure	Study Results – significant effect on progression	Study Results- no significant effect on progression
Chair stand performance				
Medial joint involvement	(28)	JSN		p = 0.944*
<i>Laboratory tests</i>				
↓Serum vitamin D	36	↑ KL grade	Lowest tertile OR 2.89 (1.01-8.25)*** p = 0.05 Middle tertile OR 2.83 (1.02-7.85)*** p= 0.05	
CPPD (in aspirated synovial fluid)	33	Radiographic features ↑ global symptoms ↑ pain	OR 2.41 (1.33-4.39)*** OR 1.89 (1.06-3.38)*** OR 1.88 (1.17-3.16)***	
Uric acid concentration	40	↓ exercise tolerance JSW	OR 1.85 (1.04-3.29)***	43-54 (mg/l) OR 1.05 (0.36-3.00)** > 54 (mg/l) OR 1.36 (0.46-4.02)**
Results are ordered by number of studies investigating characteristic first; then by strength of findings and quality of studies. †high quality studies in bold, low quality studies in parentheses; BMI = body mass index; JSW = joint space width; JSN = joint space narrowing; ↑KL grade = increase in Kellgren/Lawrence grade, 1+ by one grade, 2+ by two grades; OR = odds ratio; RR = relative risk; HR = hazard ratio; PP = predicted probability; *Univariate analysis – not adjusted; **univariate analysis - adjusted; *** multivariate analysis; ‡ = data not given for adjusted results or multivariate models; VAS = visual analogue scale; WOMAC = Western Ontario and McMaster Universities Osteoarthritis Index, (F) = function subscale; PF = patello femoral joint; OA = osteoarthritis; CPPD = calcium pyrophosphate crystals;				

Appendix G: Data for inter-rater reliability of the i-score

i-score – Rater 1	i-score – Rater 2	Difference: Rater 1 : Rater 2
24	21	3
22	26	4
24	26	2
25	21	4
22	24	2
14	14	0
31	22	9
28	29	1
29	32	3
21	24	3
25	26	1
29	29	0
19	24	5
22	23	1
19	21	2
22	26	4
24	25	1
17	10	7
26	26	0
21	20	1
20	17	3
28	27	1
20	20	0
21	22	1
17	27	10
22	26	4
20	19	1
24	23	1

INITIAL EXAMINATION: KNEE

Patient eligible for participation?

- Yes
- No
- Possible

If no or maybe, list reason(s):

Demographic Information

Height: _ _ _ cm

Weight: _ _ _ kg

BMI: _ _ _

General Questions

In the past year, have you had any fall including a slip or trip in which you lost your balance and landed on the floor or ground or lower level? If yes, how many?	<input type="radio"/> Yes <input type="radio"/> No _ _ _
Smoker?	<input type="radio"/> Never <input type="radio"/> Former <input type="radio"/> Current
Self Reported Activity Level (how often do you exercise?)	<input type="radio"/> Rarely or never <input type="radio"/> 1-2x / week <input type="radio"/> 3-4x / week <input type="radio"/> >4x / week

AHA Cardiovascular Screening (Not captured in database)

Resting Systolic Blood Pressure:	_ _ _
Resting Diastolic Blood Pressure:	_ _ _
Resting Heart Rate:	_ _ _

Known cardiovascular disease	<input type="radio"/> Yes	<input type="radio"/> No
Cardiovascular symptoms	<input type="radio"/> Yes	<input type="radio"/> No
Cardiovascular risk factors	<input type="radio"/> None <input type="radio"/> ≤ 2 <input type="radio"/> >2	
IDDM	<input type="radio"/> Yes	<input type="radio"/> No
Severe pulmonary disease	<input type="radio"/> Yes	<input type="radio"/> No

Assessment of Risk (From AHA/ACSM Screening Questionnaire)

A1	≤45 male, ≤55 female, °history, °symptoms, °risk factors	No further testing required	Safe to exercise moderate or vigorous level. If pulmonary disease or IDDM, monitor
A2	>45 male, >55 female, °history, °symptoms, °risk factors	No further testing required Could monitor HR, BP, RPE if exercise intensity increases	Safe to exercise moderate level If pulmonary disease or IDDM, monitor
A3	>45 male, >55 female, °history, °symptoms, ≥2 risk factors	No further testing required Could monitor HR, BP, RPE if exercise intensity increases	Safe to exercise to moderate level If pulmonary disease or IDDM, monitor
B	Known history (clinically stable) New York Heart Assoc. (NYHA) Level I or II	Further screening Exercise testing required Supervisor to monitor HR, BP and RPE until safety established and patient can self-monitor	If cleared by further screening, safe to exercise with supervision. If pulmonary disease or IDDM = Excluded
C	Known history Some symptoms (unstable/uncontrolled) NYHA Level III or IV		Excluded
D	Severe cardiovascular disease		Excluded

Is patient safe to exercise?	<input type="radio"/> No <input type="radio"/> Yes, with monitoring <input type="radio"/> Yes, no monitoring required
------------------------------	---

Clinical History

Knee pain	<input type="radio"/> No <input type="radio"/> One <input type="radio"/> Both
Which knee is the worst?	<input type="radio"/> Right <input type="radio"/> Left
Knee Pain Location (worse knee)	<input type="checkbox"/> Anterior knee <input type="checkbox"/> Posterior knee <input type="checkbox"/> Medial knee <input type="checkbox"/> Lateral knee
Morning Stiffness Knee \leq 30 minutes	<input type="radio"/> Yes <input type="radio"/> No
Duration of knee symptoms	<input type="radio"/> < 1 year <input type="radio"/> 1-2 years <input type="radio"/> >2-5 years <input type="radio"/> >5-10 years <input type="radio"/> > 10 years
Prior Knee Injuries Requiring Medical Intervention?	<input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Don't know
Hip Pain	<input type="radio"/> No <input type="radio"/> One <input type="radio"/> Both
Which hip is the worst?	<input type="radio"/> Right <input type="radio"/> Left
Hip Pain Location (worse hip)	<input type="checkbox"/> Anterior hip <input type="checkbox"/> Lateral hip <input type="checkbox"/> Posterior hip
Morning Stiffness Hip \leq 60 minutes	<input type="radio"/> Yes <input type="radio"/> No
Duration of hip symptoms	<input type="radio"/> < 1 year <input type="radio"/> 1-2 years <input type="radio"/> >2-5 years <input type="radio"/> >5-10 years <input type="radio"/> > 10 years

Prior Hip Injuries Requiring Medical Intervention?	<input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Don't know
Prior other leg injuries requiring medical intervention	<input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Don't know
Worse: hip or knee?	<input type="radio"/> Hip <input type="radio"/> Knee <input type="radio"/> Undecided

Associated Symptoms

Does your knee ever lock?	<input type="radio"/> Yes <input type="radio"/> No
If Y associated with pain?	<input type="radio"/> Yes <input type="radio"/> No
Does your knee give way/buckle/feel unstable?	<input type="radio"/> Yes <input type="radio"/> No
If Y associated with pain?	<input type="radio"/> Yes <input type="radio"/> No
Has the knee given way / buckled / shifted, in the last 2 days?	<input type="radio"/> I do not have the symptom <input type="radio"/> I have the symptom but it does not affect my activity <input type="radio"/> The symptom affects my activity slightly <input type="radio"/> The symptom affects my activity moderately <input type="radio"/> The symptom affects my activity severely <input type="radio"/> The symptom prevents me from all daily activities
Does your knee swell?	<input type="radio"/> Always – constant <input type="radio"/> Always – fluctuates <input type="radio"/> Intermittent <input type="radio"/> Never / haven't noticed
Does your knee get hot?	<input type="radio"/> Always – constant <input type="radio"/> Always – fluctuates <input type="radio"/> Intermittent <input type="radio"/> Never / haven't noticed

Irritability

For your most aggravating activity/position/posture, how long until symptoms worsen?	State activity: <input type="radio"/> Immediate to <1 minute <input type="radio"/> <10 minutes <input type="radio"/> <60 minutes <input type="radio"/> >1 hour																						
Rate your pain with that activity (NPRS):																							
<table border="1" style="width: 100%; text-align: center;"> <tr> <td><input type="radio"/></td> </tr> <tr> <td>0</td> <td>1</td> <td>2</td> <td>3</td> <td>4</td> <td>5</td> <td>6</td> <td>7</td> <td>8</td> <td>9</td> <td>10</td> </tr> </table>		<input type="radio"/>	0	1	2	3	4	5	6	7	8	9	10										
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>													
0	1	2	3	4	5	6	7	8	9	10													
After cessation of activity/position/posture, how long till pain returns to resting level?	<input type="radio"/> Immediate to <1 minute <input type="radio"/> <10 minutes <input type="radio"/> <60 minutes <input type="radio"/> >1 hour																						

Night Pain

How many times per night does your pain wake you?	<input type="radio"/> None <input type="radio"/> 1-2 <input type="radio"/> 3-4 <input type="radio"/> 5 or more
How many nights per week (in the last week?) is your sleep disturbed?	<input type="radio"/> None <input type="radio"/> 1-2 <input type="radio"/> 3-4 <input type="radio"/> 5 or more <input type="radio"/> Every night

Previous Treatment

Have you ever completed a physiotherapy programme for your knee?	<input type="radio"/> Yes <input type="radio"/> No
Did it help?	<input type="radio"/> Yes <input type="radio"/> No

Physical Examination**Gait Analysis**

Primary device for ambulation	<input type="radio"/> None <input type="radio"/> 1 stick or crutch <input type="radio"/> 2 sticks or crutches <input type="radio"/> Rolling walker
-------------------------------	---

Timed Up and Go

Use of Walking aid?	<input type="radio"/> Yes <input type="radio"/> No																						
Time (sec)	__ __ __ sec																						
Borg Perceived Exertion	-- --																						
Pain:																							
<table border="1" style="width: 100%; text-align: center;"> <tr> <td><input type="radio"/></td> </tr> <tr> <td>0</td> <td>1</td> <td>2</td> <td>3</td> <td>4</td> <td>5</td> <td>6</td> <td>7</td> <td>8</td> <td>9</td> <td>10</td> </tr> </table>		<input type="radio"/>	0	1	2	3	4	5	6	7	8	9	10										
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>													
0	1	2	3	4	5	6	7	8	9	10													

Standing**Posture/Alignment**

Tibiofemoral bony alignment (frontal plane)	<input type="radio"/> Normal <input type="radio"/> Valgus <input type="radio"/> Varus <input type="radio"/> Unable to assess
Tibiofemoral bony alignment (sagittal plane)	<input type="radio"/> Neutral <input type="radio"/> Hyperext __ __ degrees <input type="radio"/> FFD __ __ degrees
Involved LAA	<input type="radio"/> Low (<130) <input type="radio"/> Normal (130 – 150) <input type="radio"/> High (>150)
Uninvolved LAA	<input type="radio"/> Low (<130) <input type="radio"/> Normal (130 – 150) <input type="radio"/> High (>150)

*Lumbar Flexion ROM**Lumbar Extension ROM*

ROM	— — — degrees
Back pain?	<input type="radio"/> Yes <input type="radio"/> No

ROM	— — — degrees
Back pain?	<input type="radio"/> Yes <input type="radio"/> No

Check for aberrant motion/Gower's Sign
Check for limitation of lateral flexion

40m Self-Paced Walk

Use of Walking aid?	<input type="radio"/> Yes <input type="radio"/> No
Time (sec)	— — — sec
Borg Perceived Exertion	— —

Pain:										
<input type="radio"/>										
0	1	2	3	4	5	6	7	8	9	10

SUPINE

Posture/Alignment

Patella position	<input type="radio"/> Neutral <input type="radio"/> Medial displacement <input type="radio"/> Lateral displacement
------------------	--

Knee Flexion ROM

Involved Side	Concordant symptoms?	<input type="radio"/> Yes <input type="radio"/> No
	ROM	— — — degrees
	End feel	<input type="radio"/> Normal <input type="radio"/> Bony block <input type="radio"/> Firm <input type="radio"/> Springy
	P/R Sequence	<input type="radio"/> Normal <input type="radio"/> P-R <input type="radio"/> R-P <input type="radio"/> P=R

Uninvolved Side	Concordant symptoms?	<input type="radio"/> Yes <input type="radio"/> No
	ROM	— — — degrees
	End feel	<input type="radio"/> Normal <input type="radio"/> Bony block <input type="radio"/> Firm <input type="radio"/> Springy
	P/R Sequence	<input type="radio"/> Normal <input type="radio"/> P-R <input type="radio"/> R-P <input type="radio"/> P=R

Knee Extension ROM

Involved Side	Concordant symptoms?	<input type="radio"/> Yes <input type="radio"/> No
	ROM	— — — degrees
	End feel	<input type="radio"/> Normal <input type="radio"/> Bony block <input type="radio"/> Firm <input type="radio"/> Springy
	P/R Sequence	<input type="radio"/> Normal <input type="radio"/> P-R <input type="radio"/> R-P <input type="radio"/> P=R
Uninvolved Side	Concordant symptoms?	<input type="radio"/> Yes <input type="radio"/> No
	ROM	— — — degrees
	End feel	<input type="radio"/> Normal <input type="radio"/> Bony block <input type="radio"/> Firm <input type="radio"/> Springy
	P/R Sequence	<input type="radio"/> Normal <input type="radio"/> P-R <input type="radio"/> R-P <input type="radio"/> P=R

Ankle Dorsiflexion (in knee extension for muscle length)

Involved Side	ROM (from plantargrade)	__ __ __ degrees
---------------	-------------------------	------------------

Ankle Plantarflexion Manual Muscle testing

Involved Side	__ __ kg
---------------	----------

Hip Flexion ROM

Involved Side	ROM	__ __ __ degrees
---------------	-----	------------------

Hip Flexion Manual Muscle Testing

Involved Side	__ __ kg
---------------	----------

Hip Abduction Manual Muscle Test

Involved Side	__ __ kg
---------------	----------

Hip Adduction Manual Muscle Testing

Involved Side	__ __ kg
---------------	----------

Hamstring Flexibility

Involved Side	Concordant knee symptoms?	<input type="radio"/> Yes <input type="radio"/> No
	ROM (knee)	__ __ __ degrees

Indirect distraction of hip

Involved side: Resting Pain:

<input type="radio"/>										
0	1	2	3	4	5	6	7	8	9	10

Involved side: Within Test Pain:

<input type="radio"/>										
0	1	2	3	4	5	6	7	8	9	10

Involved side: Post Test Pain:

<input type="radio"/>										
0	1	2	3	4	5	6	7	8	9	10

X 1 Repetition SLR Strength Test

Involved Side	<input type="radio"/> Completed no quads lag <input type="radio"/> Completed with quads lag <input type="radio"/> Unable to perform
Uninvolved Side	<input type="radio"/> Completed no quads lag <input type="radio"/> Completed with quads lag <input type="radio"/> Unable to perform

Ligamentous Tests for Knee Stability

	Involved Side
MCL stress @ 30 degrees	<input type="radio"/> Normal <input type="radio"/> Grade 1 <input type="radio"/> Grade 2 <input type="radio"/> Grade 3
LCL stress @ 30 degrees	<input type="radio"/> Normal <input type="radio"/> Grade 1 <input type="radio"/> Grade 2 <input type="radio"/> Grade 3
Meniscus / McMurrays	<input type="radio"/> +ve <input type="radio"/> -ve

Accessory Movements

NB: N = Normal

P = Pain

R = Stiff

P+R = Pain + Stiff

Joint	Accessory Movement	Limit to movement			
Tib/fem	AP	<input type="radio"/> N	<input type="radio"/> P	<input type="radio"/> R	<input type="radio"/> P+R
	PA	<input type="radio"/> N	<input type="radio"/> P	<input type="radio"/> R	<input type="radio"/> P+R
Pat/fem	Med glide	<input type="radio"/> N	<input type="radio"/> P	<input type="radio"/> R	<input type="radio"/> P+R
	Lat glide	<input type="radio"/> N	<input type="radio"/> P	<input type="radio"/> R	<input type="radio"/> P+R
	Caudad	<input type="radio"/> N	<input type="radio"/> P	<input type="radio"/> R	<input type="radio"/> P+R
	Cephalad	<input type="radio"/> N	<input type="radio"/> P	<input type="radio"/> R	<input type="radio"/> P+R
	Pain on Compression	<input type="radio"/> Yes <input type="radio"/> No			
Talocrural	AP	<input type="radio"/> N	<input type="radio"/> P	<input type="radio"/> R	<input type="radio"/> P+R
	PA	<input type="radio"/> N	<input type="radio"/> P	<input type="radio"/> R	<input type="radio"/> P+R

Palpation

	Ipsilateral
Tender tib/fem medial/lateral joint lines	<input type="radio"/> Yes <input type="radio"/> No
Tender patellar margins/retropatella	<input type="radio"/> Yes <input type="radio"/> No
Tender greater trochanter	<input type="radio"/> Yes <input type="radio"/> No
Heat on Palpation	<input type="radio"/> Yes <input type="radio"/> No
Bony enlargement joint margins	<input type="radio"/> Yes <input type="radio"/> No
Bulge/Wipe test	<input type="radio"/> +ve <input type="radio"/> -ve
Patella tap test	<input type="radio"/> +ve <input type="radio"/> -ve
Crepitus with knee movement	<input type="radio"/> Yes <input type="radio"/> No

Check 1st MTP ext.*Thomas Test – Iliopsoas Flexibility*

Involved Side	ROM (hip)	__ __ __ degrees
---------------	-----------	------------------

Thomas Test – Rectus Femoris Flexibility

Involved Side	ROM (knee)	__ __ __ degrees
---------------	------------	------------------

30 second Sit-to-Stand

Use of Walking aid?	<input type="radio"/> Yes <input type="radio"/> No
No. complete	__ __
Borg Perceived Exertion	__ __

Pain:

<input type="radio"/>										
0	1	2	3	4	5	6	7	8	9	10

Side lying

Accessory Movements

NB: N = Normal

P = Pain

R = Stiff

P+R = Pain + Stiff

Joint	Accessory Movement	Limit to movement			
Sup Tib/fib	AP	<input type="radio"/> N	<input type="radio"/> P	<input type="radio"/> R	<input type="radio"/> P+R
	PA	<input type="radio"/> N	<input type="radio"/> P	<input type="radio"/> R	<input type="radio"/> P+R

Prone

Hip Internal Rotation ROM

Involved Side	Hip pain	<input type="radio"/> Yes <input type="radio"/> No
	ROM	__ __ __ degrees

Hip External Rotation ROM

Involved Side	ROM	__ __ __ degrees
---------------	-----	------------------

Knee Flexion Manual Muscle Test

Involved Side	@ 0 - 20° flx	__ __ kg
	@ 70 - 90° flx	__ __ kg
Non-Involved	@ 0 - 20° flx	__ __ kg
	@ 70 - 90° flx	__ __ kg

Gluteus Maximus Manual Muscle

Involved Side	__ __ __ kg	Check PA hypomobility.
---------------	-------------	------------------------

SITTING

Knee Extension Manual Muscle Test

Involved Side	@ 0 - 20° flx	__ __ kg
	@ 70 - 90° flx	__ __ kg
Non-Involved	@ 0 - 20° flx	__ __ kg
	@ 70 - 90° flx	__ __ kg

Medial Rotators of the Hip Manual Muscle Test

Involved Side	__ __ __ kg
---------------	-------------

Lateral Rotators of the Hip Manual Muscle Test

Involved Side	__ __ __ kg
---------------	-------------

Internal rotation knee - Passive ROM

Involved Side	_____ kg __ __ __	_____ degrees __ __ __
Involved Side	_____ kg __ __ __	_____ degrees __ __ __

External Rotation Knee – Passive ROM

Involved Side	_____ kg _ _ _	_____ degrees _ _ _
Involved Side	_____ kg _ _ _	_____ degrees _ _ _

Sock test, and Step On Stool Test

No. of steps	_ _
Borg Perceived Exertion	_ _

Pain:

<input type="radio"/>										
0	1	2	3	4	5	6	7	8	9	10

Inclusion criteria - Hip

Hip Pain + <115° hip flexion + <15° hip IR	<input type="radio"/> Yes <input type="radio"/> No
--	---

Hip pain with internal rotation + ≤ 60 minutes morning stiffness + Age 50+	<input type="radio"/> Yes <input type="radio"/> No
--	---

Inclusion criteria - Knee

Knee Pain AND three of the following: <ul style="list-style-type: none"> ● Age > 50 ● ≤ 30 minutes morning stiffness ● Crepitus ● Bony tenderness ● Bony enlargement ● No palpable warmth 	<input type="radio"/> Yes <input type="radio"/> No
--	---

Exclusion criteria

Neurogenic Disorder	<input type="radio"/> Yes	<input type="radio"/> No
Uncontrolled Hypertension	<input type="radio"/> Yes	<input type="radio"/> No
Rheumatoid Arthritis	<input type="radio"/> Yes	<input type="radio"/> No
Severe Osteoporosis	<input type="radio"/> Yes	<input type="radio"/> No
Poor Vision	<input type="radio"/> Yes	<input type="radio"/> No
Primary debilitating LBP	<input type="radio"/> Yes	<input type="radio"/> No
Previous Joint Surgery to hip or knee (excludes arthroscopy of knee)	<input type="radio"/> Yes	<input type="radio"/> No
Previous Hip Fracture	<input type="radio"/> Yes	<input type="radio"/> No <input type="radio"/> Don't know
Other surgery to LE in Previous 6 months	<input type="radio"/> Yes	<input type="radio"/> No
Current Oral Steroid Use	<input type="radio"/> Yes	<input type="radio"/> No
Steroid or Analgesic Injection prior 3 months	<input type="radio"/> Yes	<input type="radio"/> No
Recent initiation opioid analgesia/new arthritic drug (<30 days)	<input type="radio"/> Yes	<input type="radio"/> No
Weight > 155kg	<input type="radio"/> Yes	<input type="radio"/> No
Unable to ambulate > 10 m without AD	<input type="radio"/> Yes	<input type="radio"/> No
Stated inability to complete the proposed course of intervention and follow-up	<input type="radio"/> Yes	<input type="radio"/> No
Insufficient language skills to complete assessment tools	<input type="radio"/> Yes	<input type="radio"/> No
Insufficient comprehension to comply with interventions	<input type="radio"/> Yes	<input type="radio"/> No

Radiographic Information

Radiographic Grade	<input type="radio"/> None <input type="radio"/> Mild <input type="radio"/> Moderate <input type="radio"/> Severe
Osteophytes	<input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Not stated
Number of compartments of knee involved	<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> Not stated

Safety**Physical Impairments preventing safe participation?**

<p>Would this patient be at risk of injury from study interventions?</p> <p>State impairments & concerns below:</p>	<input type="radio"/> Yes <input type="radio"/> No
---	---

Prognosis

Overall assessment of prognosis	<input type="radio"/> Very good <input type="radio"/> Good <input type="radio"/> Poor <input type="radio"/> Very Poor
---------------------------------	--

Lx Spine - ↓ ROM flx/ext & hypomobile PA	<input type="radio"/> Yes	<input type="radio"/> No
Have 2 of the following: Hip/groin pain/paraesthesia, Anterior thigh pain, Passive knee flx <122°, Passive hip med rot <17°, Pain with hip distraction	<input type="radio"/> Yes	<input type="radio"/> No
Lateral knee pain including superior Tib/fib jt	<input type="radio"/> Yes	<input type="radio"/> No
Ankle d/flx - ↓ ROM (<10° by visual estimation & hypomobile AP on TC jt)	<input type="radio"/> Yes	<input type="radio"/> No
Passive 1 st MTP jt ext < 60 degrees	<input type="radio"/> Yes	<input type="radio"/> No
Abnormal Thomas test?	<input type="radio"/> Yes	<input type="radio"/> No
	If yes, consider stretching ex.	
Visual observation of trunk side flexion limitation/tightness	<input type="radio"/> Yes	<input type="radio"/> No
	If yes, consider stretching ex.	
Aberrant lumbar movement incl. catching, painful arc of motion or Gower's sign?	<input type="radio"/> Yes	<input type="radio"/> No
	If yes, consider core stability ex.	

22 November 2007

Dr Haxby Abbott
University of Otago
NZ Centre for Physiotherapy Research
School of Physiotherapy
PO Box 56
Dunedin

Dear Haxby,

Project Key: LRS/07/11/044

Full Title: Decreasing pain, disability, waiting lists and costs of osteoarthritis: Economic analysis of physiotherapy care to reduce costs of osteoarthritis.

Investigators: Dr. J Haxby Abbott, Professor G. David Baxter, Professor A. John Campbell, Associate Professor M Clare Robertson, Associate Professor Jean-Claude Theis.

Localities: Dunedin Hospital Orthopaedic Outpatient Clinic, The Centre for Physiotherapy Research and the School of Physiotherapy Clinics, University of Otago.

The above study has been given ethical approval by the **Lower South Regional Ethics Committee**. A list of members of this committee is attached.

Approved Documents

Questionnaire V2 received 22 November 2007

Certification

The Committee is satisfied that this study is not being conducted principally for the benefit of the manufacturer or distributor of the medicine or item in respect of which the trial is being carried out.

Accreditation

The Committee involved in the approval of this study is accredited by the Health Research Council and is constituted and operates in accordance with the Operational Standard for Ethics Committees, April 2006.

Progress Reports

The study is approved until **31 January 2013**. The Committee will review the approved application annually and notify the Principal Investigator if it withdraws approval. It is the Principal Investigator's responsibility to forward a progress report covering all sites prior to ethical review of the project in **November 2008**. The report form is available on <http://www.newhealth.govt.nz/ethicscommittees>. Please note that failure to provide a progress report may result in the withdrawal of ethical approval. A final report is also required at the conclusion of the study.

Requirements for SAE Reporting

The Principal Investigator will inform the Committee as soon as possible of the following:

- Any related study in another country that has stopped due to serious or unexpected adverse events
- withdrawal from the market for any reason

- all serious adverse events occurring during the study in New Zealand which result in the investigator breaking the blinding code at the time of the SAE or which result in hospitalisation or death.
- all serious adverse events occurring during the study worldwide which are considered related to the study medicine. Where there is a data safety monitoring board in place, serious adverse events occurring outside New Zealand may be reported quarterly.

All SAE reports must be signed by the Principal Investigator and include a comment on whether he/she considers there are any ethical issues relating to this study continuing due to this adverse event. It is assumed by signing the report, the Principal Investigator has undertaken to ensure that all New Zealand investigators are made aware of the event.

Amendments

All amendments to the study must be advised to the Committee prior to their implementation, except in the case where immediate implementation is required for reasons of safety. In such cases the Committee must be notified as soon as possible of the change.

Please quote the above ethics committee reference number in all correspondence.

The Principal Investigator is responsible for advising any other study sites of approvals and all other correspondence with the Ethics Committee.

It should be noted that Ethics Committee approval does not imply any resource commitment or administrative facilitation by any healthcare provider within whose facility the research is to be carried out. Where applicable, authority for this must be obtained separately from the appropriate manager within the organisation.

Yours Sincerely,



Riria Tautau-Grant
Ethics Committee Administrator
Lower South Regional Ethics Committee
email: riria_tautau-grant@moh.govt.nz

Timed Up and Go

Use of Walking aid?	<input type="radio"/> Yes
Time (sec)	<input type="radio"/> No
Borg Perceived Exertion	— — — sec — —

Pain:

<input type="radio"/>										
0	1	2	3	4	5	6	7	8	9	10

Standing

Posture/Alignment

Tibiofemoral bony alignment (frontal plane)	<input type="radio"/> Normal <input type="radio"/> Valgus <input type="radio"/> Varus <input type="radio"/> Unable to assess
Tibiofemoral bony alignment (sagittal plane)	<input type="radio"/> Neutral <input type="radio"/> Hyperext _ _ degrees <input type="radio"/> FFD _ _ degrees
Trendelenburg Stance	<input type="radio"/> None <input type="radio"/> Left <input type="radio"/> Right <input type="radio"/> Bilateral
Concordant Symptoms Painful Side?	<input type="radio"/> Yes <input type="radio"/> No
Involved LAA	<input type="radio"/> Low (<130) <input type="radio"/> Normal (130 – 150) <input type="radio"/> High (>150)

Lumbar Flexion ROM

ROM	— — — degrees
Back pain?	<input type="radio"/> Yes <input type="radio"/> No

Lumbar Extension ROM

ROM	— — — degrees
Back pain?	<input type="radio"/> Yes <input type="radio"/> No

40m Self-Paced Walk

Use of Walking aid?	<input type="radio"/> Yes <input type="radio"/> No
Time (sec)	— — — sec
Borg Perceived Exertion	— —

Pain:

<input type="radio"/>										
0	1	2	3	4	5	6	7	8	9	10

SUPINE

Posture/Alignment

Patella position	<input type="radio"/> Neutral <input type="radio"/> Medial displacement <input type="radio"/> Lateral displacement
------------------	--

Hip Flexion ROM

Involved Side	Concordant symptoms? ROM	<input type="radio"/> Yes <input type="radio"/> No — — — degrees
---------------	-----------------------------	--

Knee Flexion ROM

Involved side	End feel	<input type="radio"/> Normal <input type="radio"/> Bony block <input type="radio"/> Firm <input type="radio"/> Springy
	P/R Sequence	<input type="radio"/> Normal <input type="radio"/> P-R <input type="radio"/> R-P <input type="radio"/> P=R

Knee Extension ROM

Involved Side	End feel	<input type="radio"/> Normal <input type="radio"/> Bony block <input type="radio"/> Firm <input type="radio"/> Springy
	P/R Sequence	<input type="radio"/> Normal <input type="radio"/> P-R <input type="radio"/> R-P <input type="radio"/> P=R

Ankle Plantarflexion/ Muscle testing

Involved Side	__ __ kg
---------------	----------

Hip Flexion Muscle Testing

Involved Side	__ __ kg
---------------	----------

Hip Abduction Muscle Test

Involved Side	__ __ kg
---------------	----------

Hip Adduction Muscle Testing

Involved Side	__ __ kg
---------------	----------

Hamstring Flexibility

Involved Side	Concordant knee symptoms?	<input type="radio"/> Yes <input type="radio"/> No
	ROM (knee)	__ __ __ degrees

FABER Test

Involved Side	Positive Test?	<input type="radio"/> Yes <input type="radio"/> No
---------------	----------------	---

FADIR Test

Involved Side	Concordant Symptoms?	<input type="radio"/> Yes <input type="radio"/> No
	Cross midline?	<input type="radio"/> Yes <input type="radio"/> No

Accessory Movements

NB: N = Normal

P = Pain

R = Stiff

P+R = Pain + Stiff

Joint	Accessory Movement	Limit to movement			
Tib/fem	AP	<input type="radio"/> N	<input type="radio"/> P	<input type="radio"/> R	<input type="radio"/> P+R
	PA	<input type="radio"/> N	<input type="radio"/> P	<input type="radio"/> R	<input type="radio"/> P+R
Pat/fem	Med glide	<input type="radio"/> N	<input type="radio"/> P	<input type="radio"/> R	<input type="radio"/> P+R
	Lat glide	<input type="radio"/> N	<input type="radio"/> P	<input type="radio"/> R	<input type="radio"/> P+R
	Caudad	<input type="radio"/> N	<input type="radio"/> P	<input type="radio"/> R	<input type="radio"/> P+R
	Cephalad	<input type="radio"/> N	<input type="radio"/> P	<input type="radio"/> R	<input type="radio"/> P+R
	Pain on Compression	<input type="radio"/> Yes <input type="radio"/> No			
Talocrural	AP	<input type="radio"/> N	<input type="radio"/> P	<input type="radio"/> R	<input type="radio"/> P+R
	PA	<input type="radio"/> N	<input type="radio"/> P	<input type="radio"/> R	<input type="radio"/> P+R

Palpation

	Ipsilateral
Tender tib/fem medial/lateral joint lines	<input type="radio"/> Yes <input type="radio"/> No
Tender patellar margins/retropatella	<input type="radio"/> Yes <input type="radio"/> No
Tender greater trochanter	<input type="radio"/> Yes <input type="radio"/> No
Heat on Palpation	<input type="radio"/> Yes <input type="radio"/> No
Bony enlargement joint margins	<input type="radio"/> Yes <input type="radio"/> No
Bulge/Wipe test	<input type="radio"/> +ve <input type="radio"/> -ve
Patella tap test	<input type="radio"/> +ve <input type="radio"/> -ve
Crepitus with knee movement	<input type="radio"/> Yes <input type="radio"/> No

Thomas Test – Iliopsoas Flexibility

Involved Side	ROM (hip)	__ __ __ degrees
---------------	-----------	------------------

Thomas Test – Rectus Femoris Flexibility

Involved Side	ROM (knee)	__ __ __ degrees
---------------	------------	------------------

30 second Sit-to-Stand

Use of Walking aid?	<input type="radio"/> Yes
No. complete	<input type="radio"/> No
Borg Perceived Exertion	— — — —

Pain:

<input type="radio"/>										
0	1	2	3	4	5	6	7	8	9	10

Side lying

Accessory Movements

NB: N = Normal

P = Pain

R = Stiff

P+R = Pain + Stiff

Joint	Accessory Movement	Limit to movement			
Sup Tib/fib	AP	<input type="radio"/> N	<input type="radio"/> P	<input type="radio"/> R	<input type="radio"/> P+R
	PA	<input type="radio"/> N	<input type="radio"/> P	<input type="radio"/> R	<input type="radio"/> P+R

Prone

Hip Extension ROM

Involved Side	Concordant symptoms? ROM	<input type="radio"/> Yes <input type="radio"/> No — — — degrees
---------------	-----------------------------	--

Knee Flexion Manual Muscle Test

Involved Side	@ 0 - 20° flx	— — kg
	@ 70 - 90° flx	— — kg

Gluteus Maximus Manual Muscle

Involved Side	— — — kg
---------------	----------

SITTING

Knee Extension Manual Muscle Test

Involved Side	@ 0 - 20° flx	— — kg
	@ 70 - 90° flx	— — kg

Medial Rotators of the Hip Manual Muscle Test

Involved Side	— — — kg
---------------	----------

Lateral Rotators of the Hip Manual Muscle Test

Involved Side	— — — kg
---------------	----------

Sock Test

Grabs toes with ease	<input type="radio"/> 0
Grabs toes with effort	<input type="radio"/> 1
Beyond malleoli but not toes	<input type="radio"/> 2
Can hardly if at all reach the malleoli	<input type="radio"/> 3

Step On Stool Test

No. of steps	— —
Borg Perceived Exertion	— —

Pain:

<input type="radio"/>										
0	1	2	3	4	5	6	7	8	9	10

Manual of Operating Procedures – Baseline Knee Assessment

Inclusion Criteria

American College of Rheumatology (ACR) clinical criteria for classification of knee osteoarthritis (OA) (Altman et al., 1986).

Knee pain plus three of the following clinical features:-

- Age > 50
- Tenderness on palpation
- No heat on palpation
- < 30 min morning stiffness
- Bony enlargement
- Crepitus

To meet the primary requirement for knee pain, participants are asked if they *ever* experience knee pain.

There is no age restriction for inclusion but participants under 50 years of age must experience knee pain and fulfil three other criteria.

The margins of the medial and lateral tibio-femoral joints and the patello-femoral articulation are palpated to detect joint tenderness or bony enlargement.

Joint warmth is detected by palpation and comparison with the contralateral knee.

Morning stiffness is defined as stiffness that eases within 30 minutes of rising, and is additional to their usual restriction of knee movement.

Crepitus is confirmed by palpation of the knee joint during active flexion and extension of the knee while the participant is supine. If there is doubt about the presence of crepitus the participant repeats the test by rising from sitting. A positive response is palpable crepitus.

Exclusion Criteria

Presence of any of the following conditions excludes the individual from the trial:

- Neurological conditions (e.g. stroke, Parkinsons disease, Multiple Sclerosis)
- Uncontrolled hypertension – requiring frequent visits to doctor, or changing medications
- Rheumatoid Arthritis
- Severe osteoporosis - >1 trivial fracture (in last 2 years), or diagnosis from bone scan, or on medication (Fosamax)
- Poor vision – participant would be at risk moving around clinic
- Debilitating low back pain – severely limits functional activity. Sciatic symptoms plus positive findings on neurological testing (reflexes/ power/ sensation/ Babinski/ clonus)
- Previous joint surgery to hip or knee; excludes arthroscopy or menisectomy
- Previous hip fracture
- Other lower extremity surgery in previous 6 months
- Steroid injection - < 3 months. (can delay assessment and treatment if necessary, to fall outside 3 months)
- Oral steroids – previous long course >30 days, or current intake >5mg
- Recent initiation of opioid analgesia or new anti-arthritic drug (<30 days)
- Weight > 155kg (limit of equipment)
- Unable to ambulate > 10 m without walking aid
- Resident > 45 min from Dunedin
- Stated inability to complete course of treatment or follow-up

- Lack of comprehension/understanding. Participant does not have capacity to understand questionnaires (language or cognition)

Equipment

Long arm goniometer.

Use for: knee range of movement (ROM), Hip ROM

Knee landmarks: axis of movement is lateral femoral epicondyle, proximal arm aligned with greater trochanter and parallel with midline aspect of lateral thigh, distal arm aligned with lateral malleolus of ankle and parallel with midline aspect of lateral lower leg.

Hip landmarks: axis of movement is greater trochanter, proximal arm is aligned parallel to midline aspect of trunk, distal arm is aligned with lateral femoral epicondyle and parallel with midline aspect of lateral thigh.

Short arm goniometer

Use for: ankle ROM, Longitudinal Arch Angle of the foot (LAA)

Ankle landmarks, axis of movement is lateral malleolus, proximal arm is aligned with the fibula head parallel to lateral midline aspect of lower leg, distal arm is aligned parallel with 5th metatarsal.

Bubble inclinometer

Use for: lumbar spine ROM, hip internal/external ROM

Lumbar spine landmark is T12, located by palpating along the bottom rib.

Nicholas MMT, Hand Held Dynamometer (HHD)

Use for: all muscle strength.

The same instruction is given to every participant,

“I want you to match my pressure and build up to your maximum effort over 4 seconds. Do not allow the limb to move.”

The examiner will record the force achieved at 4 seconds or sooner if the participant moves excessively in response to resistance or stops pushing. Verbal encouragement and motivation should be used by the examiner to ensure a maximal effort by the participant. Unless otherwise stated the joint is positioned in mid range.

Other equipment: – tape measure, stopwatch, standard height chair (19cm at seat, with arms), standard height chair (18cm at seat, without arms), 20cms step, height measure, weighing scales

General Questions

- Record height and weight. Calculate BMI later. Calculate as weight (kilograms) divided by height (metres) squared and recorded as kg/m².
- Self reported activity level. “Exercise” defined as any activity completed with the intention of exercising rather than usual daily activity.
- Cardiovascular screening. Blood pressure (BP) / Heart Rate (HR) taken using the A & D Medical Digital BP monitor (Model UA – 767plus). Taken at rest on participants left arm, with the cuff at heart height and the arm relaxed. Severe pulmonary disease includes any unstable respiratory problem requiring multiple trips to medical practitioner, and poorly controlled with medication. Also any respiratory problem that would affect participants’ ability to exercise safely.
- Exercise testing will involve recording HR, BP, Rate of Perceived exertion (RPE) and participant appearance with a 6 minute bicycle

ergometer sub-maximal test. Baseline measurements are taken prior to test commencement, between 2 and 3 minutes of the test and post test. Participant appearance and RPE is monitored regularly during test. Resistance is set at a minimal level so the subject is working against a low load. The test will be terminated if the HR reaches 85% of age adjusted maximum HR or if the HR fails to rise in response to exercise. The test will be terminated if the BP exceeds 220/105 mmHg or if there is a diastolic drop of greater than 10mmHg. The participant may elect to terminate the test if RPE exceeds their personal threshold for continuing. The examiner will terminate the test if they are concerned about the subject's appearance or reported RPE. If the test is terminated or the post exercise recordings exceed the stated levels the subject will be excluded from the trial due to safety concerns. If satisfactory response occurs the participant will be included in the trial. The treating physiotherapist will be notified of the need for exercise monitoring until they have established a safe response and the ability of the participant to self-monitor.

- Exercise test – acceptable alternatives. 6min walk test on treadmill or on 10m measured walk.

Clinical History

- Variables relate to index knee and include: concordant pain with knee movement; pain report with physical performance measures; night pain; pain with locking or giving way; pain location; bilateral or unilateral knee pain; ipsilateral hip pain; the numeric pain rating scale (NPRS) (Jensen et al., 1999). The NPRS consists of three questions asking the participant to rate their pain on a scale of 0-10, where 0 is no pain at all and 10 is the worse pain imaginable, for current pain, best

pain and worst pain in the last 24 hour period (Jensen et al., 1999). The three scores are summed to give a score out of 30, with higher scores representing worse pain. This composite measure has good psychometric properties (Jensen 1999)

- Morning stiffness is defined as stiffness in the knee/hip joint on waking and rising in the morning. This is additional or different to usual joint stiffness and should ease when participant up and moving around within a designated time period (30 minutes or less for the knee, 60 minutes or less for the hip).

YES = stiffness present and eases within 30 mins.

NO = no stiffness or stiffness that does not ease within 30 mins.

- Prior injuries requiring medical intervention, defined as those which necessitated assessment and/or treatment by specialists at a tertiary facility (hospital).

Associated Symptoms

- Locking - defined as inability to bend or straighten knee for a period of time or requiring a physical manoeuvre to get knee moving.
- Giving way/buckling/instability defined as any collapse of the knee joint or feeling that it might collapse or feeling of instability of the knee.
- A standardised question from a question from the Knee Outcome Survey-Activities of Daily Living Scale (KOS-ADLS)(Irrgang et al., 1998) asks participants to categorise the impact of instability on every day activities.

Physical Assessment.

Standing

- Posture/Alignment - Tibiofemoral alignment (frontal plane) is assessed visually. The participant is asked to adopt a standing posture with legs as close together as possible.

Valgus = knees are closer together than ankles (>5 degrees visual assessment)

Normal = knees have slight valgus (<5 degrees visual assessment)

Varus = knees are further apart than ankles or are completely straight (Malanga and Nadler, 2006).

- Tibiofemoral alignment (sagittal plane) is measured with long arm goniometer (see knee use of goniometer)

Hyperextension is recorded as a positive value.

Neutral is 0 degrees.

Fixed flexion deformity is recorded as a lack of full extension and will be a negative value.

- LAA = longitudinal arch angle. The participant is asked to stand in tandem standing on elevated platforms. The measured foot is anterior with the medial aspect of the foot aligned with the edge of the platform. The participant should put equal weight through each leg. Bony landmarks are located and marked, the midpoint of the medial malleolus, the tuberosity of navicular and the medial aspect of the first metatarsal head. The short arm goniometer is located with the fulcrum being over the tuberosity of navicular. The arms are aligned with the other marked points and the examiner reads the goniometer.

Low arch = < 130 degrees

Normal = 130 – 150 degrees

High arch => > 150 degrees (McPoil and Cornwall, 2005)

- Lumbar Spine ROM Flexion - The participant is standing. The inclinometer is positioned at T12/L1. The participant is asked to bend forwards as far as possible keeping their knees straight. They are asked to report any lumbar pain reproduction (Binkley et al., 1995)
- Lumbar Spine ROM Extension - The participant is standing. The inclinometer is positioned at T12/L1. The participant is asked to arch backwards as far as possible keeping their knees straight. They are asked to report any lumbar pain reproduction (Binkley et al., 1995).

Supine

- Patella Alignment. With the knee in neutral the medial and lateral borders of the patella are palpated. The distance between the two borders is measured and the halfway point determined. The distance from the midpoint of the patella to the medial epicondyle of the femur is measured. The distance from the midpoint of the patella to the lateral epicondyle of the femur is measured (Herrington, 2000).
Lateral displacement = medial > lateral
Neutral = Medial = lateral
Medial displacement = Medial < lateral
- Knee flexion ROM. Participant positioned supine. Knee actively flexed and then passively moved to EOR. The participant or assistant can hold the knee at EOR while the examiner reads off the goniometer. Participant reports concordant symptoms. Bony landmarks of the greater trochanter, the lateral femoral condyle and the lateral maleolus to ensure correct and consistent alignment (Brosseau et al., 2001, Lin et al., 2001).

- Knee extension ROM. Participant positioned supine. The heel may be placed on a folded towel to allow any hyperextension to occur. The knee is taken passively to EOR. The examiner or assistant may hold the knee at EOR while the examiner reads off the goniometer. Concordant symptoms recorded. Bony landmarks of the greater trochanter, the lateral femoral condyle and the lateral malleolus to ensure correct and consistent alignment (Brosseau et al., 2001, Lin et al., 2001).
- Joint End Feel. The examiner also makes an assessment of end feel of joint ROM. They move the joint passively to EOR and apply overpressure if pain allows. The examiner may use the uninvolved knee for comparison.

Normal = a small amount of give at EOR. Tissue approximation in flexion.

Bony block = a very hard/bony end feel.

Firm = a very stiff end feel without a bony block.(capsular)

Springy = excessive give and is often short of EOR. If pain is limiting ROM the end feel will be described as springy as there is still more than normal give in the soft tissues. EMPTY was also included under Springy, meaning inability to push to EOR due to pain/anxiety or detection of muscle spasm (Hayes and Petersen, 2001).

- Pain Resistance sequence is defined as the participant response to the knee joint being moved passively to EOR.

Normal = no pain experienced.

P-R = pain comes on before resistance

R-P = resistance comes on before pain

P=R = pain and resistance come on together (Hayes and Petersen, 2001)

- Ankle dorsiflexion ROM. The participant is positioned supine, with the knee at full extension or as close as possible. The foot is passively

dorsiflexed and held in position by the examiner or an assistant. The goniometer is aligned and the value read by the examiner. Axis of movement being the lateral malleolus, the fixed arm being aligned with the shaft of the fibula using the bony landmark of the head of the fibula, and the moving arm aligned with the shaft of the 5th metatarsal (Elveru et al., 1988).

- Ankle Plantarflexion Muscle Test. The ankle is placed in mid range plantarflexion. The HHD is placed over the metatarsal phalangeal joints. The subject is instructed to push away by pointing their foot down.
- Hip flexion ROM. The subject is instructed and assisted to pull their thigh as close as possible to their chest. Goniometer is aligned as described and the examiner records ROM.
- Hip flexion Muscle Test. The participant may flex the opposite knee/hip to stabilise the pelvis. The involved hip is flexed to mid range. The HHD is placed proximal to the superior border of the patella. The participant is instructed to try to flex the hip, the examiner resists in the direction of extension and slight abduction.
- Hip abduction muscle test. The participant moves to the opposite side of the lowered plinth and places the uninvolved foot on the floor (use a step if necessary for comfort). This is to stabilise the pelvis. The involved hip is abducted to mid range. The HHD is placed on the lateral aspect of the distal thigh at the level of the superior border of the patella. The participant is instructed to push out with the involved leg. The examiner resists with a medially directed force.
- Hip adduction muscle test. The participant remains in position described. The hip is adducted to mid range. The HHD is placed on the medial aspect of the distal end of thigh in line with the superior border

of the patella. The participant is instructed to push in with the involved leg while the examiner resists with a laterally directed force

- Hamstring flexibility (90/90 test). The participant is returned to normal supine position. The uninvolved leg remains on the plinth. The involved leg is flexed to 90 degrees of hip flexion. This position may be maintained by the participant or an assistant. The examiner aligns the goniometer as per knee ROM. The examiner extends the knee until resistance in the hamstrings muscles prevents further extension. The movement is recorded in degrees from full extension. Hamstring tightness is defined as a greater than 30 degree loss of full knee extension with the femur held at 90 degrees of hip flexion (Bandy and Irion, 1994). Concordant knee symptoms are recorded.
- Hip indirect distraction. The participant is in supine. The examiner records resting hip symptoms. The ankle is cradled in the examiners hands with the hip in slight flexion/abduction/external rotation. The examiner imparts a sustained (10 secs) longitudinal force in the caudad direction. The participant is asked to report any hip symptoms during the application of the force. The hip is then returned to neutral resting position and the participant asked to report any resting symptoms. (Care needs to be taken if knee symptoms increase for a knee participant. The force may have to be applied proximal to the knee joint).
- Medial Collateral Ligament Stress test. The examiner sits on the plinth next to the subject on the side that is to be tested. They support the subject's knee in 30 degrees of flexion. The hand on the lateral side of the knee is placed over the lateral joint line. The other hand is placed on the medial aspect of the lower third of the tibia. The lateral hand imparts a medial force to the knee joint. The lower hand stabilises the

tibia or can impart a small lateral force. The pain response of the subject and the degree of movement allow the test to be graded.

Grade I = medial pain produced, normal amount of joint play

Grade II = increased amount of joint play produced but with definite end feel. Pain usually produced.

Grade III = excessive amount of joint play produced, no definite end feel. Usually no pain (Magee, 2002).

- **Lateral Collateral Ligament Stress Test.**The examiner abducts the subject's leg at the hip. The knee is supported in 30 degrees of flexion. One hand is placed on the medial aspect of the knee over the medial joint line. The other hand is placed on the lateral aspect of the lower third of the tibia. A lateral force is imparted by the hand over the medial joint line. The lower, lateral hand may stabilise the lower leg or exert a medial force to the lower tibia (Magee, 2002). The pain response of the subject and the degree of movement allow the test to be graded as above.
- **Mcmurrays Test (Dervin et al., 2001).** The involved knee is flexed to end of range. If participant is unable to tolerate end of range flexion, the tests will be performed at 90 degrees of knee flexion. External tibial rotation and axial loading is applied as the knee is extended. The test is repeated with internal tibial loading.
A positive test is when the subject reports reproduction of concurrent pain or the examiner detects a clunk or click with performance of the test.
- **X1 SLR Test –** The subject is asked to lock their knee as straight as possible. They must then lift their straight leg a short distance (20cms) off the bed keeping the knee as straight as possible. If they are unable to keep the knee locked to its full limit of extension, slight flexion

occurs and is seen as a lag. (note if the participant has a fixed flexion deformity there has to be additional loss of their available range of extension to be classified as a quads lag)

Performed no lag = SLR with no loss of knee extension

Performed with lag = leg lifted off bed but knee is seen to flex

Unable to perform = the subject is unable to lift the leg off the bed

- Accessory Movements. As the accessory movements are performed, the examiner judges the amount of joint glide detected with a normal amount of force. If the examiner has to use excessive force to produce a movement between the joint surfaces, or the degree of movement is less than expected, then resistance or stiffness (R) is noted. The examiner may compare with the contralateral joint. If the participant reports pain during the accessory movement then pain (P) is noted.

Normal = no P, no R

P = Pain only reported

R = Resistance (stiff) only detected

P+R = Pain reported and Resistance detected (stiff)

Accessory movements for the tibio-femoral joint, the patello-femoral joint and the talo-crural joint are all performed with a rolled up towel under the thigh, producing slight knee flexion.

- Tibio-femoral.
 - AP - hand placed over the anterior aspect of the knee and a posterior force applied over the tibial plateau.
 - PA - both hands are cradled around the posterior aspect of the knee joint and the tibia lifted up by exerting an anterior force through the tibial condyles. The thumbs of both hands can palpate the anterior joint line to assist with the assessment of joint glide.
- Patello-femoral.

Medial glide - thumbs or thenar eminence will be placed over the lateral border of the patella. A medial force is imparted to make the patella glide.

Lateral glide - thumbs or thenar eminence will be placed over the medial border of the patella and a lateral force applied. Care must be taken not to compress the patella onto the femur.

Caudad glide - examiner cups their hand over the patella with the thenar and hyperthenar eminences contacting the superior border of the patella. A downward (towards the feet) glide of the patella is produced with the line of force being parallel to the femur and avoiding any compression of the patello-femoral joint.

Cephalad glide - examiner cups their hand over the patella with the thenar and hyperthenar eminences contacting the inferior pole of the patella. An upwards (towards the head) glide of the patella is produced with the line of force being parallel to the femur and avoiding any compression of the patellofemoral joint.

- Patello-femoral Compression test. The examiner uses a broad contact between their hand and the patella, then produces a posterior force which compresses the patella onto the femur. A pain response is recorded (Wood et al., 2006).
- Talo-crural. The examiner allows the foot and ankle to adopt its resting position.

AP - The examiner puts the web space between their thumb and index finger over the anterior aspect of the ankle joint, holding the talus. They stabilise the tibia with their other hand. A posterior force is applied to the talus.

PA – the examiner holds the posterior aspect of the ankle joint cradling the calcaneus. They stabilise the tibia with the other hand. The

examiner exerts an anterior force moving the calcaneus and talus together and resisting movement of the tibia.

- Palpation. The knee is positioned in comfortable flexion to allow easy detection of the tibiofemoral joint lines during palpation.

The knee is supported in slight flexion for palpation of patellar margins and retropatellar surface. The patella may have to be moved to palpate the retropatellar surface.

The greater trochanter of the hip is located by palpating down the lateral side of the thigh until a bony prominence is located. The degree of tenderness can be compared to the contralateral side if participant response is equivocal.

Heat is detected by general palpation with the posterior aspect of the hand and comparison to the contralateral knee.

Bony Enlargement can be detected at the medial and lateral aspects of the tibiofemoral joint close to the joint lines. It is usually the tibial plateau or femoral condyles which appear enlarged. Also the patella may appear to be enlarged.

- Bulge/Wipe test (Magee, 2002). The anterior aspect of the knee is massaged from distal to proximal and slight lateral direction. Long sweeping strokes are used to “milk” the fluid in the joint superiorly. The lateral side of the joint is then wiped from a proximal to distal direction. Any bulge observed on the medial aspect of the joint in response to the downward wipe is recorded as a positive test.
- Patella Tap (Magee, 2002). The examiners hands encircle the patella from the superior and inferior directions. This pressure is maintained while the thumb and index finger of one hand are used to tap down on the anterior surface of the patella. A positive test is when the posterior

aspect of the patella does not contact sharply with the femur but appears to float on the underlying swelling.

- Crepitus. The participant is asked to flex and extend their knee. The examiner has their hand placed lightly over the knee joint. Any crunching/clicking/grating sounds are recorded as crepitus. The patient is positioned supine initially. If response is equivocal the manoeuvre is repeated in standing.
- Thomas Test. The participant perches on the end of the elevated plinth with their feet resting on the floor. The uninvolved thigh is flexed towards the chest. The participant maintains this position while they rock back into supine (* Note : an assistant is often required for this test). This should enable the lumbar spine to flatten onto the plinth, excessive pelvic tilt should be avoided. The involved leg is allowed to rest down. Hip extension is measured and recorded with a goniometer. If the thigh does not reach neutral (horizontal) the value recorded is negative, if it drops below horizontal it is recorded as positive (Clapis et al., 2008). The participant is encouraged to relax the knee and knee flexion from full extension is measured and recorded with a goniometer.

Side - lying

- Accessory movement superior tibio-fibular joint. The participant is positioned in side lying with their knees in 30 degrees of flexion (approx). The involved side is uppermost. A folded towel or cushion may be required if the pressure on the medial aspect of the knees is painful.

AP – the examiner stands at the side of the plinth in front of the participant. An AP force is imparted to the anterior aspect of the joint. Pain (P) and Resistance (R) is recorded as previously described.

PA – the examiner stands at the side of the plinth behind the participant and imparts a PA force to the joint. Pain (P) and Resistance (R) is recorded as previously described.

Prone

- Hip internal rotation ROM – The involved hip is positioned by lifting the thigh and placing it in neutral internal and external rotation and in neutral alignment with the trunk. The opposite hip is placed in 15 degrees of abduction. The involved knee is flexed to 90 degrees. The inclinometer is placed on the medial maleolus and zeroed. The hip is internally rotated by passively dropping the involved heel out to the side. The hip is moved to R1 which is when the pelvis starts to rise from the plinth. The ROM and concurrent pain are recorded.
- Hip external rotation ROM – The hip is positioned in neutral as above. The inclinometer is placed on the lateral maleolus and zeroed. The hip is externally rotated by passively moving the heel to the midline. The hip is moved to R1 which is when the pelvis starts to rise from the plinth. The ROM and concurrent pain are recorded.
- Knee flexion muscle tests. These are performed on the involved and then the uninvolved side. The knee is first flexed to 90 degrees or as close as possible. The HHD is placed on the posterior aspect of the ankle just above the calcaneus. The participant is instructed to pull their heel towards the buttock. The examiner resists by pushing the opposite way. The knee must stay between 70 and 90 degrees of flexion. The knee is then fully extended and the same direction of

movement attempted. The knee must stay between 0 and 20 degrees of flexion. The tests are repeated on the uninvolved leg. (* Note : the participants frequently report sensations of hamstring cramp when the test is performed at 90 degrees knee flexion. This often resolves if the test is performed @ 0-20 degrees. Alternatively a pillow can be placed under the abdomen to produce slight lumbar flexion and prevention cramp).

- Gluteus maximus muscle test. The participant flexes the involved knee to 90 degrees. The HHD is placed on the posterior aspect of thigh just proximal to knee crease. The participant is instructed to lift their foot towards the ceiling, lifting the thigh from the plinth and extending the hip. The examiner resists with a downward pressure.

Sitting

- Knee extension muscle tests. These tests are performed on the involved and then the non-involved sides. First the participant is instructed to let the knee relax to 90 degrees of knee flexion. The HHD is placed just proximal to the anterior aspect of the ankle joint. (* Note : extra padding may be required if anterior shin is tender or the skin is fragile). The examiner tucks their testing elbow into their side and stabilises themselves by holding onto the plinth. The participant is instructed to attempt to straighten their knee, the examiner resists with a posterior pressure. Movement should not exceed the position of 70 – 90 degrees of flexion (range < 20 degrees tolerated). The knee is extended to as close to full extension as possible. The examiner exerts a downward pressure while the participant resists any movement. 0 – 20 degrees of flexion is permitted (< 20 degrees tolerated). The tests are repeated with the uninvolved leg.

- Hip medial rotation muscle test. The participant remains in sitting with the knee flexed to approx 90 degrees. The examiner moves the involved hip into mid range internal rotation by bringing the ankle out to the side. The HHD is positioned on the lateral aspect of the lower leg just above the lateral maleolus. The participant is instructed to hold this position against the medial force of the examiner “don’t let me push you down”.
- Hip lateral rotation muscle test. The participant is positioned as above. The examiner moves the hip into mid range external rotation by pulling the ankle up towards midline. The HHD is placed on the medial aspect of the lower leg, just above the medial maleolus. The participant is instructed to hold this position against the lateral force of the examiner “don’t let me push you out”.

Physical Performance Measures

Time is recorded in seconds to the nearest tenth/hundredth of a second.

Pain during test is recorded in NPRS

Total effort of performing test is reported as Borg’s Rating of Perceived Exertion (RPE)

- Timed Up and Go – A standard height (19cm) armchair with armrests is used. The participant is asked to stand, using armrests if necessary, walk to the 3 metre line, cross the line, turn round, walk back, turn round and sits down. They are instructed to walk and not run, they may use a walking aid if necessary, and they should walk as quickly as possible but still safely, and keep moving. The timing commences when the participant’s buttock leaves contact with the chair. The turn at 3 metres is not included but the turn to sitting is included. The

timing stops when the participant's buttocks reconnect with chair (Podsiadlo and Richardson, 1991, Stratford and Kennedy, 2006).

- 40 metre self-paced walk – Participants walk up and down a 10 metre course 4 times. They are instructed to “walk as quickly as possible but still so you are comfortable”. They may use a walking aid if necessary. The turnaround times were excluded. Adapted from (Kennedy et al., 2005).
- 30 second sit to stand. The participant is seated on a standard height chair without arms. They are not permitted to use their hands to push up from the chair and must fold their arms across their chest. They must rise to a fully erect position with knees and hips straight and then sit down again. This complete cycle counts as one repetition. The timing starts as the buttocks leave the chair. The number of complete cycles in 30 seconds is recorded. The participant is encouraged to complete as many repetitions as possible (Guralnik et al., 1994, Jones et al., 1992).
- Step Test – A 20 cms height step is used. Use of handrails or support is not permitted. The participant is asked to place the involved foot up on the step, then step up with the uninvolved leg. The uninvolved leg is returned to the ground first followed by the involved leg. The involved leg is therefore the working leg. The number of steps is recorded. The test is performed at the participant selected speed. They are encouraged to complete as many steps as possible, up to a maximum of 50. Verbal encouragement is used by the examiner to get the best score. The participant may elect to stop due to pain, shortness of breath, weakness or lack of energy, or because they feel unsafe (loss of balance or instability of knee) (van den Dikkenberg et al., 2002, Mookink et al., 2005).

Table D Inter-rater reliability of clinical measures reported with ICC (2,1) in participants with knee or hip osteoarthritis – results for all tests performed

Test/Measure	Level Of Agreement *	ICC(2,1)	95% CI	SEM
Self-paced walk, seconds	Good	0.95	0.90, 0.98	2.00
Timed up and Go, BORG	Good	0.92	0.83, 0.96	0.72
Step to stool test, count	Good	0.91	0.82, 0.96	5.80
Step-to-stool test, BORG	Good	0.88	0.77, 0.94	1.22
Step-to-stool test, NPRS	Good	0.88	0.75, 0.94	0.90
Timed up and go, seconds	Good	0.87	0.74, 0.94	0.84
Self-paced walk, NPRS	Good	0.84	0.69, 0.92	1.03
30 second sit to stand, count	Good	0.81	0.63, 0.91	1.27
Hip flexion, ROM	Good	0.80	0.60, 0.91	8.60
Self-paced walk, BORG	Good	0.79	0.61, 0.90	1.26
Standing lumbar extension, ROM	Good	0.77	0.56, 0.88	4.50
Timed up and go, NPRS	Good	0.76	0.55, 0.88	1.16
30 second sit to stand, BORG	Moderate	0.75	0.53, 0.87	1.33
Hip flexion, strength	Moderate	0.74	0.48, 0.87	2.13
Knee extension 90 strength	Moderate	0.72	0.48, 0.86	3.51
Hip external rotation, strength	Moderate	0.72	0.48, 0.86	1.23
Thomas Test, psoas flexibility, ROM	Moderate	0.71	0.47, 0.85	6.69
Hip internal rotation, strength	Moderate	0.71	0.47, 0.85	1.28
Hip extension, ROM	Moderate	0.68	0.40, 0.84	5.44
Knee flexion 20 strength	Moderate	0.68	0.25, 0.86	2.80

Appendix L: Tables of results for inter-rater reliability investigation for all hip and knee clinical tests

Test/Measure	Level Of Agreement *	ICC(2,1)	95% CI	SEM
30 second sit to stand, NPRS	Moderate	0.66	0.39, 0.83	1.49
Hip abduction, strength	Moderate	0.65	0.38, 0.82	1.47
Thomas test, rectus flexibility, ROM	Moderate	0.60	0.31, 0.79	12.2
Hamstring flexibility, ROM	Moderate	0.58	0.28, 0.78	7.58
Standing lumbar flexion, ROM	Moderate	0.53	0.23, 0.76	10.6
Ankle plantarflexion, strength	Poor	0.40	0.02, 0.68	4.14
Knee alignment, sagittal ROM (Variance)	Poor	0.38	-0.03, 0.67	5.43
Hip extension, strength	Poor	0.37	-0.10, 0.70	3.31
Knee extension 20 strength	Poor	0.37	-0.11, 0.71	4.60
Hip adduction, strength	Poor	0.33	-0.11, 0.68	3.49
Knee flexion 90 strength	Poor	0.18	-0.20, 0.51	4.73
* level of agreement ICC > 0.75 = good, 0.5 – 0.75 = moderate, < 0.5 = poor (Portney and Watkins, 2000); ICC = intraclass correlation coefficient; 95% CI = 95% confidence interval; SEM = standard error of measurement; BORG = measure of perceived exertion; NPRS = Numerical Pain Rating Scale; ROM = range of motion				

Table E Inter-rater reliability of clinical variables reported with kappa (κ) in participants with knee or hip osteoarthritis – results of all tests performed

Test/Measure	Level of Agreement *	Kappa	95% CI	% agreement
FADIR test, concordant symptoms	Excellent	0.86	(0.67, 1.0)	93.10
Sock Test	Substantial	0.75	(0.67, 0.91)	85.71
Hip flexion, concordant symptoms	Substantial	0.66	(0.39, 0.93)	82.76
FABER test	Substantial	0.64	(0.38, 0.91)	82.14
FADIR test, cross midline	Substantial	0.63	(0.35, 0.92)	82.76
Tibiofemoral P-A	Substantial	0.63	(0.18, 1.0)	93.10
Longitudinal arch angle	Moderate	0.60	(0.35, 0.82)	86.20
Tenderness medial/lateral joint lines	Moderate	0.60	(0.34, 0.86)	79.31
Knee extension, end feel	Moderate	0.59	(0.53, 0.76)	79.31
Bony enlargement	Moderate	0.58	(0.28, 0.87)	79.31
Knee flexion, end feel	Moderate	0.56	(0.52, 0.63)	75.86
Back pain(flexion), concordant symptoms	Moderate	0.51	(0.04, 0.99)	89.66
Hip extension, concordant symptoms	Moderate	0.49	(0.19, 0.80)	75.86
Heat on palpation	Moderate	0.46	(-0.17, 1.0)	93.10
Patellar compression test	Moderate	0.44	(0.03, 0.85)	82.76
Knee flexion, pain resistance sequence	Moderate	0.41	(0.28, 0.56)	62.07
Talocrural A-P	Moderate	0.41	(0.14, 0.55)	75.00

Appendix L: Tables of results for inter-rater reliability investigation for all hip and knee clinical tests

Test/Measure	Level of Agreement *	Kappa	95% CI	% agreement
Trendelenburg test, concordant symptoms	Fair	0.40	(0.06, 0.75)	70.37
Tenderness patellar margins	Fair	0.37	(0.03, 0.71)	68.97
Superior tibiofibular P-A	Fair	0.35	(0.27, 0.57)	62.07
Knee alignment, frontal	Fair	0.30	(0.21, 0.64)	51.72
Patellar caudal glide	Fair	0.29	(0.14, 0.58)	58.62
Knee extension, pain resistance sequence	Fair	0.28	(0.20, 0.39)	58.62
Talocrural P-A	Fair	0.26	(0.15, 0.32)	57.14
Tenderness greater trochanter	Fair	0.23	(-0.08,0.54)	62.07
Back pain(extension), concordant symptoms	Fair	0.22	(-0.20,0.63)	75.86
Crepitus	Fair	0.22	(-0.13,0.57)	68.97
Superior tibiofibular A-P	Fair	0.22	(-0.12,0.48)	72.41
Knee alignment, sagittal	Fair	0.21	(-0.25,0.68)	55.17
Trendelenburg test	Slight	0.19	(0.06, 0.34)	42.86
* Level of agreement κ : 0.81 – 1 = excellent, 0.61 – 0.8 = substantial, 0.41 – 0.6 = moderate, 0.21 – 0.4 = fair, 0.0 – 0.2 = slight, < 0.0 = poor (Landis and Koch, 1977); 95% CI = 95% confidence interval; FABER = flexion, abduction, and external rotation; FADIR = flexion, adduction, and internal rotation; P-A = postero – anterior accessory movement; A-P = antero – posterior accessory movement				

Patient Global Rating of Change (GRC)

Adapted from (Jaeschke et al., 1989)

How do you rate the change in your knee symptoms since you started physiotherapy?

- +7 Very great deal better
- +6 A great deal better
- +5 Quite a bit better
- +4 Moderately better
- +3 Somewhat better
- +2 A little bit better
- +1 A tiny bit better
- 0 About the same
- 1 A tiny bit worse
- 2 A little bit worse
- 3 Somewhat worse
- 4 Moderately worse
- 5 Quite a bit worse
- 6 Great deal worse
- 7 Very great deal worse

CLINICIAN INITIAL VISIT - KNEE MANUAL THERAPY GROUP

Please rate your current knee pain
0 1 2 3 4 5 6 7 8 9 10

Best pain in past 24 hours
0 1 2 3 4 5 6 7 8 9 10

Worst pain in last 24 hours
0 1 2 3 4 5 6 7 8 9 10

Irritability **Low** **Medium** **High**

Functional squat Pain
0 1 2 3 4 5 6 7 8 9 10

ROM **degrees**

Pain with gait? **Yes** **No**

Prescription of walking aid? **Yes** **No**

Regional Screening

Lx Spine - ↓ ROM flx/ext & hypomobile PA **Yes** **No**

Have 2 of the following: Hip/groin pain/paraesthesia,
 Anterior thigh pain, Passive knee flx < 122 degrees
 Passive hip med rot < 17 degrees, Pain with hip distraction **Yes** **No**

Lateral knee pain including superior Tib/fib jt **Yes** **No**

Ankle d/flx - ↓ ROM (<10 degrees by visual estimation
 & hypomobile AP on TC jt) **Yes** **No**

Passive 1st MTP jt ext < 60 degrees **Yes** **No**

Appendix N: Standardised physiotherapy treatment protocols

Knee Flexion	Joint Position	Grade	Dose	Variations
AP mobilisation Tibio/Femoral joint	<input type="checkbox"/> Pain free range <input type="checkbox"/> P1/R1 range <input type="checkbox"/> EOR	<input type="checkbox"/> III <input type="checkbox"/> IV	___ sets 30 reps	
<input type="checkbox"/> Medial P/F glide <input type="checkbox"/> Lateral P/F glide <input type="checkbox"/> Caudad P/F glide <input type="checkbox"/> Cephalad P/F glide	___ degrees knee flx	<input type="checkbox"/> III <input type="checkbox"/> IV	___ sets 30 reps	
Pure physiological flx	<input type="checkbox"/> Pain free range <input type="checkbox"/> P1/R1 range <input type="checkbox"/> EOR	<input type="checkbox"/> III <input type="checkbox"/> IV	___ sets 30 reps	
Knee flexion stretch	___ degrees knee flx		2 X 60 secs	
STM Quads/Peripatellar/ITB	___ degrees knee flx		___ mins	
Knee Extension	Joint Position	Grade	Dose	Variations
PA mobilisation to Tibio/femoral joint	<input type="checkbox"/> Pain free range <input type="checkbox"/> P1/R1 range <input type="checkbox"/> EOR	<input type="checkbox"/> III <input type="checkbox"/> IV	___ sets 30 reps	
External rotation to tibio/femoral joint	___ degrees knee flx	<input type="checkbox"/> III <input type="checkbox"/> IV	___ sets 30 reps	
Pure physiological extension	<input type="checkbox"/> Pain free range <input type="checkbox"/> P1/R1 range <input type="checkbox"/> EOR	<input type="checkbox"/> III <input type="checkbox"/> IV	___ sets 30 reps	
Hamstrings stretch	___ deg hip flx		2x60sec	

Appendix N: Standardised physiotherapy treatment protocols

Gastroc stretch	___ deg knee ext			
STM Hamstrings STM Gastroc/Adductors			___ mins	

Secondary Techniques. 5 minute max treatment per impairment.

Hip Joint	Indicate if used	Reps (x30)
Long axis distraction/thrust	<input type="checkbox"/> V <input type="checkbox"/> IV <input type="checkbox"/> III	___ sets
Caudal glide progression	<input type="checkbox"/> IV <input type="checkbox"/> III	___ sets
AP Progression	<input type="checkbox"/> IV <input type="checkbox"/> III	___ sets
Medial/inferior glide	<input type="checkbox"/> IV <input type="checkbox"/> III	___ sets
PA in FABER	<input type="checkbox"/> IV <input type="checkbox"/> III	___ sets
PA progression	<input type="checkbox"/> IV <input type="checkbox"/> III	___ sets
Internal Rotn in Ext	<input type="checkbox"/> IV <input type="checkbox"/> III	___ sets
STM – quadriceps STM – Hamstrings STM – post/lat hip	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	2-3 mins each
Quad/hip flexor stretch Hamstring stretch Glute/ER stretch Hip IR stretch	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	2x60 secs each

Appendix N: Standardised physiotherapy treatment protocols

Ankle Joint	Indicate if used	Dose
Rearfoot distraction manipulation	<input type="checkbox"/> V	
AP to talo-crural jt	<input type="checkbox"/> IV <input type="checkbox"/> III	___ sets 30 reps
Physiological d/flx	<input type="checkbox"/>	___ sets 30 reps
Gastroc/Soleus STM Plantarfascia STM	<input type="checkbox"/> <input type="checkbox"/>	2 – 3 mins each
Gastroc stretch	<input type="checkbox"/>	2x60 secs each
Lumbar Spine		
Lumbopelvic manipulation	<input type="checkbox"/> V	
Lumbar rotation mobilization	<input type="checkbox"/> V <input type="checkbox"/> IV <input type="checkbox"/> III	___ sets 30 reps
SNAG flx/ext	<input type="checkbox"/>	___ reps
STM to paravertebral muscles	<input type="checkbox"/>	2 – 3 mins
Superior Tibio/Fibula Joint		
AP mobilisation	<input type="checkbox"/> IV <input type="checkbox"/> III	___ sets 30 reps
PA mobilisation	<input type="checkbox"/> IV <input type="checkbox"/> III	___ sets 30 reps
STM to popliteal fossa/ITB	<input type="checkbox"/>	2-3 mins
1st MTP Joint		
Dorsal Glide to 1 st MTP	<input type="checkbox"/> IV <input type="checkbox"/> III	___ sets 30 reps
Physiological extension 1 st MTP	<input type="checkbox"/> IV <input type="checkbox"/> III	___ sets 30 reps
STM to plantarfascia	<input type="checkbox"/>	2-3 mins
1 st MTP extension stretch	<input type="checkbox"/>	2x60 secs

Post Treatment Assessment

Please rate your current knee pain ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○
 0 1 2 3 4 5 6 7 8 9 10

Functional squat Pain ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○
 0 1 2 3 4 5 6 7 8 9 10

ROM ___ ___ ___ **degrees**

Time taken to complete treatment ___ ___ **mins**

_____date

_____Therapist Signature

CLINICIAN INITIAL ASSESSMENT FORM

KNEE EXERCISE GROUP

Please rate your current (hip or knee) pain

0 1 2 3 4 5 6 7 8 9 10

Best (least) pain in the past 24h

0 1 2 3 4 5 6 7 8 9 10

Worst pain in the past 24h

0 1 2 3 4 5 6 7 8 9 10

Functional Squat

Pain Level

0 1 2 3 4 5 6 7 8 9 10

ROM — — —

Pain with gait? Yes No

Did you prescribe an assistive device? Yes No

EXAMINATION FOR PRESCRIPTION OF ADDITIONAL EXERCISES

Inability to perform 10 bilateral heel raise repetitions to FULL height with FULL heel inversion? Yes No
↳if yes, include heel raise progression

Inability to perform 10 side-lying hip abduction repetitions? Yes No
↳if yes, include abduction progression

Inability to perform 10 side-lying hip external rotation repetitions? Yes No
↳if yes, include ext. rotation progression

Abnormal Thomas test? Yes No

↳if yes, consider stretching exercises

Visual observation of trunk side flexion limitation/tightness Yes No
↳if yes, consider stretching exercises

Aberrant lumbar movement including catching, painful arc of motion, or Gower's sign? Yes No
↳if yes, consider core stability progression

Appendix N: Standardised physiotherapy treatment protocols

Knee Techniques	Intervention	Tick if used	Level of intervention (Please tick)	# of sets completed	
Aerobic (up to 10 min)	Cycling or Treadmill			___ ___ min	
Strengthening (3x10)	Knee extensors 1		<input type="radio"/> 1, <input type="radio"/> 2, <input type="radio"/> 3		
	Knee extensors 2		<input type="radio"/> 1, <input type="radio"/> 2, <input type="radio"/> 3		
	Knee fl/ Hip ext.		<input type="radio"/> 1, <input type="radio"/> 2, <input type="radio"/> 3		
	Knee flexors		<input type="radio"/> 1, <input type="radio"/> 2, <input type="radio"/> 3		
Stretching (2x60 sec)	Calf muscle group				
	Hamstring m group.				
	Quadriceps m group				
Neuromuscular Control (2 min each, 6 min total)	Weight Shifts		<input type="radio"/> 1, <input type="radio"/> 2, <input type="radio"/> 3		
	Balance		<input type="radio"/> 1, <input type="radio"/> 2, <input type="radio"/> 3		
	Braiding		<input type="radio"/> 1, <input type="radio"/> 2, <input type="radio"/> 3		
	Stairs				
	Shuttle walking				
Additional Techniques	5 minutes maximum per impairment				
▪ Weak ankle plantarflexors	Heel raises		<input type="radio"/> 1, <input type="radio"/> 2, <input type="radio"/> 3		
▪ Weak hip abductors	Hip abduction		<input type="radio"/> 1, <input type="radio"/> 2, <input type="radio"/> 3		
▪ Weak hip ext. rotators	Hip ext. rotation		<input type="radio"/> 1, <input type="radio"/> 2, <input type="radio"/> 3		
▪ Thomas test positive	Stretching IP/RF/TFL				
▪ Weak core stability	Abdominal Bracing		<input type="radio"/> 1, <input type="radio"/> 2, <input type="radio"/> 3		
▪ Trunk side flexion tightness	Side Flexion Stretch				
Post Treatment Pain Level?	<div style="text-align: right;"> <input type="radio"/> </div> <div style="text-align: right;">0 1 2 3 4 5 6 7 8 9 10</div>				
Post Treatment Functional Squat	Pain Level?	<div style="text-align: right;"> <input type="radio"/> </div> <div style="text-align: right;">0 1 2 3 4 5 6 7 8 9 10</div>			
	ROM?	___ ___ ___			
Total time treated (min)	___ ___ min.				

Appendix N: Standardised physiotherapy treatment protocols

Appendix O: WOMAC and GRC baseline and one-year data

		Baseline WOMAC Scores			One year GRC Scores		One Year WOMAC Scores		
Patient ID No.	Rx Group	Pain	Function	Total	GRC Score	GRC Responder	Pain	Function	Total
26230	0	60	57.65	59.58	13	1	12	18.24	17.08
31195	0	40	41.76	43.75	5.5	0	30	27.06	28.75
44262	0	18	21.76	21.67	7	0	10	10.59	11.67
91466	0	50	37.06	42.08	7	0	38	24.12	28.33
124316	0	38	29.41	32.5	8	0	46	38.24	39.58
162365	0	12	11.18	12.92	6.5	0	12	19.41	19.17
184421	0	24	34.71	32.08	7	0	42	28.82	31.67
241254	0	68	59.41	62.08	12	1	54	45.29	47.5
259206	0	24	48.82	45.42	6.5	0	30	64.12	59.17
279344	0	12	20.59	20.42	7	0	22	18.24	19.58
311351	0	46	61.18	57.92	5.5	0	64	71.76	70.42
344198	0	20	25.29	23.33	7	0	4	7.06	6.25
374401	0	18	11.76	15.42	11	1	12	13.53	12.92
466432	0	0	1.18	1.25	7.5	0	2	1.76	2.92
541412	0	16	34.12	30.42	2	0	54	67.06	62.08
641332	0	42	56.47	53.75	10	1	28	38.82	38.75
643366	0	36	25.29	26.67	7	0	26	28.24	25.42
646265	0	54	58.82	57.08	2	0	42	44.12	43.75
659284	0	68	71.76	71.67	14	1	2	9.41	9.17
660400	0	52	43.53	48.75	7	0	54	44.12	48.75
765190	0	26	26.47	27.5	4.5	0	30	27.06	29.58
786314	0	96	75.29	80.83	7	0	80	89.41	83.33
822259	0	34	45.88	45.42	8	0	30	48.82	45
842356	0	10	14.71	13.75	8	0	0	12.94	9.17
877476	0	22.5	14.71	18.85	7	0	14	6.47	9.58
907148	0	16	21.76	20.42	6	0	24	30.59	28.75
992370	0	22	4.12	10	6.5	0	28	20	23.75
993237	0	62	61.76	62.5	1.5	0	86	80.59	82.5
24219	1	6	15.29	12.92	12	1	2	5.29	4.17
28367	1	16	19.41	19.58	11.5	1	2	31.76	23.75
65409	1	20	68.82	59.58	6	0	52	54.71	55.42
98398	1	42	43.53	45.42	1.5	0	56	61.76	62.5
140287	1	52	46.47	48.75	12	1	24	20.59	24.17
150298	1	20	11.18	14.58	12.5	1	2	1.18	1.67

Appendix O: WOMAC and GRC baseline and one-year data

		Baseline WOMAC			One year GRC		One Year WOMAC		
		Scores			Scores		Scores		
168410	1	26	21.18	22.08	7	0	18	7.65	9.17
183233	1	34	35.88	37.08	4	0	72	52.94	58.33
215328	1	4	0	1.25	8.5	0	6	0.59	2.08
262501	1	32	18.24	23.75	11	1	20	31.18	30.42
337329	1	20	26.47	25.42	9	0	16	21.76	20.42
391188	1	72	77.65	76.67	12	1	12	21.76	19.58
392204	1	60	46.47	50.42	6.5	0	76	67.06	70
435364	1	42	44.71	44.58	2	0	34	37.65	37.5
444391	1	8	2.35	3.75	10	1	4	5.29	5
479294	1	40	45.88	45.42	13.5	1	4	4.12	4.17
529446	1	52	67.65	63.33	10	1	28	28.82	28.75
587315	1	44	20.59	27.08	4.5	0	34	23.53	25.42
609306	1	40	35.88	37.08	12	1	34	25.88	25.42
627434	1	42	34.71	37.08	14	1	18	18.24	18.75
649439	1	48	24.71	32.08	9.5	0	30	26.47	27.5
694256	1	58	63.53	60.83	6	0	46	44.71	45
747382	1	28	34.12	36.67	6	0	44	33.53	37.08
833442	1	10	13.53	14.58	7	0	4	10.59	10.42
861217	1	82	82.94	83.33	13.5	1	56	41.18	45.42
897436	1	32	29.41	34.17	13.5	1	20	5.29	9.58
986266	1	82	74.71	76.67	3	0	48	48.24	47.08
994224	1	10	14.12	13.33	14	1	8	11.18	11.25
49407	2	0	2.35	3.75	8.5	0	2	5.88	5.42
92308	2	56	51.18	53.75	12.5	1	10	2.94	5.42
103271	2	18	37.65	30.42	9	0	36	34.71	33.75
109209	2	14	4.71	9.58	7	0	36	32.35	32.92
135465	2	82	87.65	86.67	11.5	1	30	34.12	33.33
137291	2	84	80	81.25	10	1	32	44.12	40.42
169430	2	20	22.35	20	13.5	1	8	4.12	4.58
241338	2	46	44.12	46.25	12	1	6	11.18	10.83
298456	2	54	71.18	67.08	5.5	0	66	68.24	62.08
371244	2	20	14.12	15	13	1	8	7.65	7.5
417251	2	58	51.76	52.92	12.5	1	32	32.35	33.33
429431	2	68	73.53	70.83	2	0	60	80	72.92
464281	2	82	78.24	80	13	1	4	5.29	5.42
466258	2	40	38.82	41.25	12.5	1	20	8.24	14.17

Appendix O: WOMAC and GRC baseline and one-year data

		Baseline WOMAC			One year GRC		One Year WOMAC		
		Scores			Scores		Scores		
515320	2	2	2.35	2.5	10	1	0	0	0
563343	2	46	46.47	45.83	5	0	60	67.06	65.42
585385	2	52	26.47	33.75	10	1	10	8.82	10
607376	2	52	58.24	57.08	5.5	0	70	70.59	70.42
711236	2	56	60.59	61.25	5.5	0	36	55.29	50
740354	2	10	31.76	27.92	3.5	0	30	42.35	42.92
762358	2	12	2.35	5.83	9.5	0	2	0.59	0.83
770380	2	14	22.94	22.92	14	1	0	12.94	11.25
826214	2	72	68.24	70	10	1	30	14.12	17.92
837269	2	92	93.53	92.92	10.5	1	30	71.18	59.58
856272	2	54	64.71	63.75	12.5	1	22	31.18	30.42
888416	2	32	28.24	30	9	0	22	25.29	26.25
900201	2	82	84.12	83.75	12	1	32	35.88	34.58
962406	2	54	38.82	42.5	10	1	18	16.47	17.92
998423	2	50	23.53	28.33	7	0	44	10.59	17.08
25282	3	28	35.88	34.58	6	0	40	28.24	31.67
62429	3	18	20.59	21.67	6.5	0	30	34.71	36.25
90260	3	62	75.29	74.17	6	0	16	53.53	43.33
93302	3	48	38.24	43.33	9.5	0	8	26.47	22.92
151327	3	62	53.53	55.83	13	1	0	26.47	22.5
157321	3	8	5.88	6.25	7.5	0	4	5.29	5
207451	3	14	7.06	7.92	7	0	4	11.18	8.75
251194	3	44	27.65	32.08	6	0	50	22.94	27.92
269426	3	34	37.06	35.83	7	0	18	20.59	19.17
271373	3	12	0.59	2.92	7.5	0	12	1.18	4.17
303369	3	56	55.88	56.67	6	0	54	41.76	47.08
341453	3	40	40.59	41.25	9	0	12	25.29	22.08
355277	3	68	60.59	63.33	9	0	64	63.53	63.75
356247	3	40	35.29	37.92	14	1	10	5.29	5.83
389239	3	64	55.88	60	8	0	38	31.18	35
410404	3	18	27.06	29.17	5.5	0	26	24.71	23.75
413384	3	44	69.41	62.5	10	1	76	37.65	44.17
500471	3	46	41.18	44.58	4.5	0	50	55.29	56.25
554241	3	62	58.82	60.42	10.5	1	36	47.65	45
651231	3	60	58.82	58.75	9	0	50	43.53	47.08
663399	3	22	9.41	11.25	6.5	0	10	8.82	9.58

Appendix O: WOMAC and GRC baseline and one-year data

		Baseline WOMAC Scores			One year GRC Scores		One Year WOMAC Scores		
700353	3	30	48.24	42.08	2	0	38	45.29	44.17
731245	3	74	77.65	76.25	5.5	0	40	62.94	54.17
735383	3	38	41.18	43.75	13	1	6	4.71	7.5
767492	3	28	11.18	18.33	4	0	26	15.29	19.58
839200	3	72	75.88	75.83	7	0	50	41.76	44.17
935296	3	20	8.82	12.5	14	1	0	0	0
977449	3	42	23.53	30.42	8.5	0	24	18.24	20.42
<p>Key: WOMAC = Western Ontario and McMaster Universities osteoarthritis index, converted to maximum score of 100; GRC = Global Rating of Change; ID No. = participant identification number; Rx group = treatment group: 0 = usual care, 1 = exercise therapy only, 2 = manual therapy only, 3 = exercise and manual therapy</p>									

Appendix P: OARSI responder data

WOMAC Pain Response 50%/ 20+	WOMAC Function Response 50%/ 20+	OARSI Responder Pain or Function* 50%/ 20+	WOMAC Pain Response 20%/ 10+	WOMAC Function Response 20%/ 10+	GRC Response† >10	OARSI Responder Pain, Function, GRC‡ 2 out of 3	OARSI Responder at One Year§
0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1
0	0	0	0	0	0	0	0
1	0	1	1	1	1	1	1
0	0	0	1	0	1	1	1
0	0	0	1	0	0	0	0
1	1	1	1	1	1	1	1
0	1	1	1	1	1	1	0
0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0
0	1	1	1	1	1	1	0
1	1	1	1	1	1	1	0
0	0	0	0	0	0	0	0
1	0	1	1	1	0	1	0
1	1	1	1	1	1	1	1
0	0	0	1	0	0	0	0
0	0	0	1	0	1	1	1
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0
0	0	0	0	0	1	0	0
Total non-responders per category							
88	93	85	55	71	62	63	78
Total responders per category							
25	20	28	58	42	51	50	35
Key: WOMAC pain/function response = change in pain or function subscales of WOMAC; 50%/20+ = change of 50% or change of more than 20 raw score (/100); *OARSI responder = 50%/20+ change in pain OR function; 20%/10+ = change of 20% or change of more than 10 raw score (/100); †GRC score of 10 or greater (“somewhat better”); ‡ OARSI responder = 2 out of 3 of pain/function/GRC (20%/10+); §OARSI responder at one year = primary outcome = either 50%/20+ responder OR 2 out of 3 responder; 1 = yes; 0 = no							