Modelling clinical outcomes and cost-effectiveness of primary care interventions for osteoarthritis using prediction and decision models

Jerome Wulff

A thesis submitted for the degree of Doctor of Philosophy

June, 2013

Arthritis Research UK Primary Care Centre

Keele University

Declaration

This project is lodged within the North Staffordshire Osteoarthritis Projects (NorStOP) funded by a programme grant awarded by the Medical Research Council (MRC) UK (grant code G9900220) and the National Institute of Health Research (NIHR) UK (grant code RP-PG-0407-1038). The proposal for this study represents the first theme of the NIHR OA programme grant titled "Modelling optimal primary care for OA" acquired by the Arthritis Research UK Primary Care Centre in 2008.

The planning, design and ethical submissions of the NorstOP cohorts was undertaken by a team of researchers led by Professor Peter Croft, Professor George Peat, Professor Krysia Dziedzic and Dr John McBeth whilst the participants in the NorStOP studies were identified and recruited by the network and population survey teams and data entry and quality checks were performed by the administrative team all at the Centre.

I received guidance and advice from my three supervisors Dr Milisa Bucknall, Professor Danielle van der Windt and Dr Sue Jowett on the outline of the thesis, statistical and economic analyses, interpretation and writing of the chapters and discussion. Dr Pelham Barton advised me on the more precise aspects of economic modelling study and I also received guidance from Dr Nadia Corp on the search strategy for the systematic review and meta-analysis I carried out.

I performed all the analyses of the prediction modelling study, evidence synthesis and meta-analysis study and the decision modelling of economic evaluation study.

Acknowledgments

I would like express my profound thanks to my supervisors Dr Milisa Bucknall, Professor Danielle van der Windt and Dr Sue Jowett for their immeasurable advice and support throughout the development of this thesis. I also thank the Arthritis Research UK Primary Care Centre for offering me the studentship to undertake this project.

I would like to extend my gratitude to the National Institute of Health Research (NIHR) UK for funding my project and the Medical Research Council (MRC) UK for funding the North Staffordshire Osteoarthritis Projects (NorStOP) data I used.

I am grateful to the coordinators of the NorStOP cohorts for making their data available for me to use, the network and population survey teams for identifying and recruiting participants in the NorStOP cohorts and the administrative team for carrying out the data entry and quality checks of the NorStOP participants.

My gratitude also goes to all the OA clinicians and experts particularly Prof Peter Croft, Prof George Peat, Prof Krysia Dziedzic, Prof Nadine Foster, Dr Mark Porcheret and Dr John Bedson for their valuable advice on the design and structure of both the prediction and economic modelling studies.

I would like to thank Dr Pelham Barton of the Health Economics Unit at the University Birmingham for his valuable advice during the development of the decision model. I am also grateful to Prof Michael Hurley and Dr Nicola Walsh of Kings College, London for sending me the usual care cost data of their ESCAPE trial study which I used for my decision model study.

I am also thankful to Dr Nadia Corp for offering me guidance on the search strategy for my evidence synthesis and meta-analysis study, Dr Elaine Thomas for happily providing me with the numerous data I requested, Dr Martyn Lewis for sending and explaining the algorithm for deriving utility scores to me, Dr Kelvin Jordan for sending me the paper on adjusted PAR and Mr Daniel Green for estimating the cost of medication in the NorStOP consultation dataset.

I would also like to say thank you to all the staff and students at the Arthritis Research UK Primary Care Centre (Keele University) for their support. I also want to thank a good friend Guy Woolnough for his continuous encouragement over the years as well as proof reading my work.

Before I end, I am grateful to my family for their tolerance and endurance during the development of this thesis.

Above all, I wish to thank Yeshua Al-Mashia for granting me strength throughout the process and for the future.

Abstract

The overall aim of this thesis is to develop prediction models to identify key predictors of poor outcome in people with osteoarthritis (OA) and examine the cost effectiveness of two approaches to delivering primary care interventions for OA compared to current primary care.

This thesis is comprised of two parts – the first part concerns the development of prediction models to identify the combination of factors that predicts poor outcome of OA in relation to pain and functional limitation at three year follow up for participants aged 50 years or more. The strongest baseline predictors of pain and functional disability were having pain in the previous year and poor physical function at baseline respectively. The models developed showed good internal validity and hence may be further tested for external validity in community-based adults with similar characteristics as those in this study.

The second part involves a summary of evidence on the effectiveness of four primary care interventions (information and advice, simple analgesia, topical NSAIDs and exercise) in reducing pain and improving function at one or more joint sites among osteoarthritis patients in primary care. The results showed significant small to moderate improvements in pain and functional disability for advice/information, topical NSAIDs and exercise interventions compared to their controls, whilst simple analgesia failed to demonstrate significant improvements in either measures. This evidence was used to populate the economic (decision) model developed in this thesis.

The decision model examined the cost effectiveness of two approaches to delivering primary care interventions for OA - stepped care and one-stop-shop care were compared with current primary care. The primary results were robust to changes in the input

variables with stepped care emerging as the most cost-effective option ahead of one-stop-shop care and current care in that order.

These findings need to be confirmed in samples of primary care consulters.

Table of Content

List of tables

List of graphs

List of abbreviations

ACR	-	American College of Rheumatology
AIC	-	Akaike Information Criteria
AIHW	-	Australian Institute of Health and Welfare
AIMS	-	Arthritis Impact Measurement Scale
ARI	-	Absolute Risk Increase
ASMP	-	Arthritis Self-Management Program
AU $	-	Australian Dollar
AUSCAN	-	Australian/Canadian Osteoarthritis Hand Index
BCa	-	Bias Corrected and Accelerated
BEEP	-	Benefits of Effective Exercise for knee Pain
BIC	-	Bayesian Information Criteria
BMI	-	Body Mass Index
BNF	-	British National Formulary
BSSNI	-	Berkman-Syme Social Network Index
CBA	-	Cost Benefit Analysis
CDN $	-	Canadian Dollar
CEA	-	Cost Effectiveness Analysis
CI	-	Confidence Interval
CINAHL	-	Cumulative Index to Nursing and Allied Health Literature
CMA	-	Cost Minimization Analysis
COX-2	-	Cyclooxygenase 2 inhibitors
CUA	-	Cost Utility Analysis
DIY	-	Do It Yourself
EO	-	Expected Optimism
EQ-5D	-	EuroQol 5 Dimensions
ES	-	Evidence Synthesis
EULAR	-	European League against Rheumatism
FAST	-	Fitness Arthritis and Seniors Trial Scale
FE	-	Fixed Effect
FPDI	-	Foot Pain Disability Index

GARS - Groningen Activity Restriction Scale
GISCA - Italian Group of Study of the Costs of Arthritis
GLM - Generalized Linear Model
GNP - Gross National Product
GoF - Goodness of Fit
GP - General Practitioner
HAD - Hospital Anxiety and Depression
HAQ - Health Assessment Questionnaire
HS - Health Survey
ICER - Incremental Cost Effectiveness Ratios
IPCHS - Institute of Primary Care and Health Sciences
IPQ - Illness Perceptions Questions
IRGL - Influence of Rheumatic Diseases on General Health and Lifestyle
IRR - Incidence Rate Ratio
ISPOR - International Society for Pharmaco-economics and Outcomes Research
JSW - Joint Space Wide
K & L - Kellgren and Lawrence
KAP - Keele Assessment of Participation
LR - Likelihood Ratio
MA - Meta Analysis
Man - Body Manikin
MAR - Missing At Random
MCAR - Missing Completely At Random
McGill Q - McGill Questionnaire
MeSH - Main Medical Subject Heading
MI - Multiple Imputation
MICE - Multiple Imputation by Chained Equation
MLE - Maximum Likelihood Estimation
MOSAIC - Management for OSteoArthritis in ConsultationS study
NCC-CC - National Collaborating Centre for Chronic Conditions
NHANES - National Health and Nutrition Examination Survey
NHS - National Health Service
NICE - National Institute for Health and Clinical Excellence

NMAR	-	Not Missing At Random
NNB	-	Number Needed to Benefit
NNH	-	Number Needed to Harm
NNT	-	Numbers Needed-to-Treat
NorStOP	-	North Staffordshire Osteoarthritis Project
NRS	-	Numerical Rating Scale
NSAIDs	-	Non Steroidal Anti-Inflammatory Drugs
NS-LREC	-	North Staffordshire Local Research Ethics Committee
OA	-	Osteoarthritis
OCP	-	Optimism Corrected Performance
OR	-	Odds Ratio
PAR	-	Population Attributable Risks
PhD	-	Doctor of Philosophy
PSA	-	Probabilistic Sensitivity Analysis
PSSRU	-	Personal Social Services Research Unit
QALY	-	Quality Adjusted Life Year
QoL	-	Quality of Life
RA	-	Rheumatoid Arthritis
RCT	-	Randomized Control Trial
RE	-	Random Effects
RoB	-	Risk of Bias
ROC	-	Receiver Operating Characteristic curve
ROM	-	Range of Motion
RPS	-	Regional Pain Survey
RR	-	Relative Risk
SD	-	Standard Deviation
SE	-	Standard Error
SF12	-	Short Form 12 Health Survey Questionnaire
SF36	-	Short Form 36 Health Survey Questionnaire
SF-6D	-	Short Form Health Questionnaire 6 Dimensions
SIP	-	Sickness Impact Profile
SMD	-	Standardized Mean Difference

SR	-	Systematic Review
SSE	-	Sum of Squares Errors
TP	-	Test Performance
UK	-	United Kingdom
USA	-	United States of America
VAS	-	Visual Analogue Scale
VIF	-	Variance Inflation Factor
WHO	-	World Health Organization
WOMAC	-	Western Ontario and McMaster Universities Osteoarthritis Index

Chapter One

Thesis overview

This chapter provides a general background of this thesis, introduces the aims and objectives, and provides an outline of each of the chapters. The thesis is structured in two parts: Part 1 describes the prediction modelling for osteoarthritis (OA) study, covered in chapters 3 to 5; Part 2A describes the evidence synthesis and meta-analysis of results regarding the effectiveness of primary care interventions for OA study, presented in chapters 6 and 7; Part 2B describes the design of a health economic decision model for the primary care management of OA study and is covered in chapters 8 and 9 whilst chapter 10 provides the general discussion and conclusions for the whole thesis. Finally, the summary of the content of each chapter is outlined at the end of this chapter.

1.1 Background and rationale

Osteoarthritis (OA) is one of the most common joint disorders in middle aged and older people [*Brooks 2002; Reginster 2002; Leigh et al 2001*] with about 8.5 million people suffering from the clinical syndrome of OA [*Arthritis Care 2004*] in the UK. It is the most common cause of mobility limitations in older people in most developed countries [*Gupta et al 2005; Thomas et al 2004a; Buckwalter et al 2004*].

Although any joint in the body can be involved, OA most commonly affects the hands, knees, hips, and feet. The precise pathophysiology of OA is not well understood but it is characterised by both loss of articular cartilage [*Your Total Health 2008; NICE 2008*] as well as metabolically active repair processes including remodelling of adjacent bone and new bone formation [*NICE 2008; Grainger and Cicuttini 2004*]. Once a person develops OA particularly in the knee the condition often gets worse over time and can lead to severe

pain and disability. Given that OA is a long-term condition, it impacts on physical, mental and social functioning, and there are multiple options for treatment where successful patient management requires a holistic (optimal) patient-centred approach [*NICE 2008*]. If successful management is not achieved, this may lead to significant consequences and costs for patients, households, health care systems and the nation as a whole [*Reginster 2002; Elders 2000; Lapsley et al 2001*].

OA is usually diagnosed clinically by identifying symptoms and signs such as pain, stiffness, tenderness, limitation of movement and/or occasional swelling of a joint [*Murphy et al 2008; Lawrence et al 2008*]. The symptoms and signs of OA vary but pain and functional disability are the main consequences of OA and can cause either short or long-term difficulty depending on the severity of the condition and the number of joints affected. Most research has focused on regional pain (i.e. one joint site) even though most people have generalized OA which is characterised by the involvement of two or more joints or groups of joints [*Gunther et al 1998*].

Current treatments particularly pharmaceutical drugs such as Non-Steroidal Anti-Inflammatory Drugs (NSAIDs) and other analgesics designed to relieve pain have considerable risk of side effects, which increases the morbidity and mortality associated with OA [*Akarca 2005; Lewis et al 2002; Emkey et al 2004*]. As a result, the National Institute for Health and Clinical Excellence (NICE) guideline has recommended that non-pharmacological strategies such as advice and exercise should be combined with drugs as the preferred treatment for OA [*NICE 2008*]. The interventions classified as core treatments by NICE which should be offered to every OA patient include advice and information, exercise and interventions (such as appropriate diets) which will help

overweight and obese OA patients' lose weight. However, if further treatment is required reasonably safe pharmacological interventions such as paracetamol and topical NSAIDs should be considered before stronger drugs such as opioids, oral NSAIDs or cyclo-oxygenase 2 (COX-2) inhibitors. These interventions have been found to be effective in reducing pain and improving function in both the short and long term that is why they are recommended by NICE.

The development of statistical models to identify factors that predict poor long term outcome in people with OA may facilitate the identification of high risk groups or might give insight into factors that could be better targeted by interventions. Review of the literature shows that only a few predictive studies [*Zhang et al 2011; Sa et al 2011; Yusuf et al 2011; Thomas et al 2008; Jinks et al 2008; Mallen et al 2007; Topp et al 2000*] for OA have been carried out, but most were limited to specific joints rather than people with any joint pain and none considered developing optimal prediction models for OA regardless of the joints involved apart from *Yusuf et al 2011* whose study involved people with knee and hip OA. There is also lack of evidence in the literature regarding the estimation of attributable risks associated with these predictors, which was used to select high risk predictors as it helps to understand maximum achievable health gains if successful interventions were to be implemented to counteract the negative effects of predictors for OA. Currently, no study has developed models to explore the cost effectiveness of providing optimal primary care for OA which would involve applying the core primary care interventions recommended for patients with OA by NICE. This will enable health care systems to make informed decisions about the choice of the most cost effective strategy to deliver optimal primary care for OA.

The key aims of the study therefore are to develop optimal prediction models for estimating poor outcome (severe pain and functional limitation) of OA in a population-based sample of older people followed-up for 3 years, to estimate the attributable proportion and number needed to treat associated with these predictors and to model the cost-effectiveness of optimal primary care (delivered in step and one-stop-shop fashions) for OA.

1.2 Part 1 – Prediction modelling study

This initial part of the work is concerned with development of optimal models for prediction of poor outcome of OA and identification of key predictors to facilitate formation of subgroups at high risk of poor outcome.

The specific objectives are:

(i) To develop optimal models of OA - i.e. determine the optimal combination of factors associated with poor outcome of OA measured as severe pain and functional limitation after 3 years follow-up.

(ii) To examine the goodness-of-fit and performance of the models.

(iii) To internally validate the models.

(iv) To estimate population attributable risk (PAR) for each predictor – i.e. the maximum achievable health gain if optimal management would reduce/prevent the adverse effect of a predictor.

(v) To estimate number needed to treat (NNT) for each predictor – i.e. the number needed to treat to prevent one additional person from suffering with OA.

(vi) To use information on the strength of association of predictors with outcome and PAR and NNT estimates to identify the most important set of predictors of poor outcome of OA.

1.3 Part 2A – Evidence synthesis and meta-analysis of primary care interventions for OA

This part of the work involves carrying out evidence synthesis (ES) and meta-analysis (MA) to estimate the overall effects of core primary care interventions (namely information and advice, simple analgesics, exercise and topical non-steroidal anti-inflammatory drugs) for osteoarthritis which are recommended in the NICE OA guidelines. Interventions to lose weight for obese patients were not considered as that is not the intention or focus of this study and also the advice/information generally given to patients' covers interventions to lose weight. The effect estimates calculated for the core primary care interventions for OA were subsequently used to populate the cost effectiveness model developed in Part 2B of this thesis.

The specific objectives for this study are:

(1) To carry out an evidence synthesis by identifying relevant randomised controlled trials (RCTs) on the effectiveness of the four interventions in primary care populations.

(2) To assess the risk of bias within each of the selected RCTs.

(3) To extract relevant data from the selected RCTs to calculate effect size estimates for the four interventions.

(4) To examine evidence of heterogeneity among effect estimates and subsequently use appropriate methods to pool them.

(5) To explore the possibility of publication bias.

1.4 Part 2B – Modelling cost effectiveness of primary care interventions for OA.

Economic modelling of the cost-effectiveness of implementing evidence-based primary care interventions for OA is the focus of the final part of this thesis. It intends to estimate the cost-effectiveness of different approaches to delivering primary care (stepped care and one-stop-shop care for OA compared to usual care).

The specific objectives for the economic modelling study are:

(i) To determine the cost effectiveness of optimal care (i.e. stepped care and one-stop-shop care) for managing OA compared with usual care.

(ii) To carry out deterministic sensitivity analyses to compare the findings with that of the base case.

1.5 Outline of the thesis

This section provides an outline of the thesis, including a brief summary of chapter two to ten.

Summary of chapter two

Chapter two describes the definition, causes and the main predictors associated with OA. It also explains the use of symptomatic (clinical) and radiographic definitions of OA and summarises information on the prevalence of OA in the United Kingdom and other

developed countries. Finally, it illustrates the economic implications of OA in terms of health care costs and indirect costs relating to productivity losses.

Summary of chapter three

Chapter three describes the background and the objectives of the prediction modelling study. The study design is outlined and target population defined. Definitions of the two outcome measures used (persistent pain and functional limitation) are provided and the various predictors used for developing the models presented. The precise procedures adopted for developing and validating prediction models and the calculation of epidemiological measures are outlined with all relevant formulae given where needed.

Summary of chapter four

Results of the prediction modelling study are presented in this chapter including those obtained from Poisson and logistic regression models. The goodness of fit and performance of each model developed is evaluated. Finally, the selection of the six most important predictors of severe pain and functional limitation is illustrated, aimed to help with the identification of high risk groups.

Summary of chapter five

The conclusion and discussion of the prediction modelling study is described in chapter five. This covers the summary of the findings, comparison with existing literature, strengths and limitations of the study and conclusion.

Summary of chapter six

Chapter six summarizes the background, objectives, data extraction, quality assessment, and pooling strategies used in the evidence synthesis (ES) and meta-analysis (MA) study. The latter is aimed at obtaining the overall effect estimates of four primary care interventions for OA, namely information and advice, simple analgesia, topical non-steroidal anti-inflammatory drugs (NSAIDs) and exercise.

Summary of chapter seven

The results of the review and meta-analysis are described in chapter seven. It (1) illustrates the characteristics of the studies used to examine the risk of bias within each study, (2) calculates the pooled effect estimates of advice and information, paracetamol, topical NSAIDs and exercise used in primary care and (3) examines the existence of heterogeneity and publication bias for each intervention.

Summary of chapter eight

This chapter summarizes the definition, rationale and concerns of economic evaluation and economic modelling. It describes the interventions to be assessed in the economic model, and summarizes the types of economic evaluations and economic decision models that can be used. The objectives of the economic modelling study are outlined and all the components used to develop the economic model are described and justified.

Summary of chapter nine

Chapter nine illustrates the results and discussion of the economic modelling study presented in chapter 8. It evaluates the base case cost effectiveness estimates for stepped

care, one-stop-shop care and usual care and also describes the results of sensitivity analyses to examine the robustness of the base case findings.

Summary of chapter ten

Chapter ten includes a general discussion of the thesis as a whole, with emphasis on the main decisions taken during the developments of the studies presented in this thesis and the implications of their findings for future research

Chapter Two

General background of OA

This chapter describes the general background of osteoarthritis with particular consideration of the general definition of OA, possible causes and predictors, diagnosis, management, prevalence, and the economic impact of OA in terms of health care costs and productivity losses.

2.1 Definition of OA

Osteoarthritis (OA), which is also known as degenerative joint disease, is a multi-factorial disorder characterised by both wear and tear of the protective cap of cartilage at the ends of bones [*Your Total Health 2008; NICE 2008*] as well as remodelling and repair of joint tissues [*Grainger and Cicuttini 2004; NICE 2008; NCC-CC 2008*]. It is the leading cause of musculoskeletal disability in both developed and developing countries [*Brooks 2002; Rabenda et al 2006; NICE 2008; NCC-CC 2008*], and is characterised by pain, aching, stiffness, and bone enlargements in and around joints.

2.2 Predictors associated with onset and progression of OA

A considerable amount of research has been conducted over the decades to investigate the predictors associated with onset or progression of OA. Exposure to a predictor makes a person more susceptible to developing or progressing with a condition, but this does not necessarily imply a causal relationship: some people exposed to the predictor will not develop the disease, whereas people without any predictors may still develop the disease.

The effects of predictors for developing and progressing with OA may vary for the different joints and also among men and women or among different age groups.

Predictors linked with the development or the progression of OA can be classified as either non-modifiable (predisposed) and modifiable. Examples of non-modifiable predictors of OA are female sex, old age, genetic factors, family history and race [*Felson and Zhang 1998; Felson et al 2000; Sowers 2001*] whilst examples of modifiable predictors are injury, obesity, occupational overuse and joint misalignment [*Felson et al 1992; Felson et al 1995; Hart et al 1999; Hunter et al 2002; Sharma 2001*]. Although precise aetiological factors of OA are not known, a number of genetic [*Neame et al 2004*], patient-specific and environmental predictors [*Felson et al 2000*] have been linked to its development and progression. Patient-specific factors include obesity, female sex and increasing age, whilst environmental factors involve occupational overuse (repetitive use) of joints over long periods of time. The identification of genes responsible for osteoarthritis requires further investigation [*Doherty et al 1994, Neame et al 2004*] although some genes have been implicated in association studies for OA [*Spector and MacGregor 2004*].

The prediction modelling study described in this thesis focuses on prognostic factors (predictors of the progression of OA) and not aetiological factors (risk factors for onset of OA); this section describes both as most predictors of the onset of OA are also predictors of the progression of OA. Below is a brief description of predictors that have been linked to OA, most of which were included in the prediction modelling study presented in the first part (chapters 3 to 5) of this thesis.

Gender

It is a well-established finding that risk of onset and progression of OA is higher among women than men [*Felson and Zhang 1998; Ding et al. 2007*]. Women are affected more often, more severely and at more sites [*Ding et al. 2007*] compared to men, particularly after menopause. The effects of hormones produced by women on cartilage may vary with menopausal status and phase of osteoarthritis. Other factors such as the different distribution of weight in women compared to men, and the possible advantages of the larger bone and the more muscular body composition of men, may be some of the reasons why women have higher risk of OA than men [*Ding et al 2003*]. Also, Ding et al [2007] reported that women have a higher rate of knee cartilage tissue loss than men although the reasons for this are not known.

Age

Although OA may begin at any age, it usually affects older people. The mean age of the onset of OA is approximately 45 years [*Ledingham et al 1993b*]. Jinks et al [2008] in their population-based study of predictors of onset and progression of symptomatic knee OA in adults confirmed that old age is associated with both onset and progression of symptomatic knee OA. Also, a radiographic study [*Nuki 1998*] has showed that there is a steady rise in OA modifications/progression in joints from 30 years onward and that by the age of 65 years, approximately 80% of their population showed some radiographic evidence of OA with about a quarter reporting some pain or disability. This age related link with OA may be due to the diminished capacity for cartilage repair, hormonal changes and the cumulative effects of environmental exposures [*Peterson and Jacobssen 2002*].

Injury

People with past record of joint injury or trauma are more likely to develop or aggravate OA particularly of the knee and hip [*Wilder et al. 2002; Gelber et al. 2000*]. These injuries are normally caused by sporting or recreational activities and may include acute joint trauma resulting from dislocation, contusion, fracture, tears of the cartilage or ligaments, and surgical meniscectomy [*Englund et al. 2003; Felson et al. 2000*]. An injury that damages the tissues within the joint can increase the stress on the cartilage. This therefore initiates the process of OA developing slowly over many years before it begins to cause symptoms of pain, stiffness or problems of mobility in the previously injured joint. Although surgical repair of an initial trauma is widely undertaken and appears to reduce pain and improve function in the short to medium term, in the long-term it may not protect the joint from developing OA [*Englund et al. 2003*].

Obesity

Being overweight or obese can lead to the development or progression of OA particularly in women [*Jinks et al 2008*]. This usually happens as a result of a general increase in body weight, thereby increasing the pressure on the cartilage and ligament, which over time results in osteoarthritis of weight-bearing joints. However, OA can also occur at non-weight bearing joints such as the joints of the hand. This signifies that obesity may also have biologic properties that support OA in a non-mechanical way via altered lipid mechanisms or interactions between insulin resistance and inflammation [*Aspden et al 2001*].

Some studies have shown that obesity is strongly associated with OA of the knee [*Jinks et al 2008*]. It was reported that obesity may become a predictor of OA as early as in the third decade of life before the inception of symptoms such as pain, stiffness and functional disability [*Felson 1988*]. OA may also have a reverse relationship with excess weight in the sense that painful joints may limit physical activity thereby causing weight gain.

Occupational overuse

People in occupations that require repetitive use of their joints are at increased risk of onset and progression of OA of the knee and hip. Jobs involving continuous repetitive activities such as kneeling, squatting and climbing stairs are associated with an increased risk of osteoarthritis of the knee, while jobs such as heavy lifting including farming and construction are associated with increased risk of osteoarthritis of the hip [*Schouten et al. 2002; Lau et al. 2000*]. It was further reported that agricultural workers with 10 years active work experience have double the risk of developing OA compared with those with less than 1 year work experience. Frequent minor injuries at joints that are continuously used either through sports, leisure or occupational activities may lead to rapid progression of OA and hence result in worsening of symptoms that can take longer time to manage. Moreover, Felson and Zhang [1998] reported that almost 30% of all knee OA is attributable to occupational activity, which involves repeated knee bending, kneeling, squatting, or climbing.

History of pain and functional disability

Dawson et al [2005] found that previous pain score and previous number of painful knee or hip joints were strongly associated with knee and hip OA progression in older adults in a general population sample. Others have also found that baseline hip pain was associated

with the progression of hip OA in samples of older adults [*Ledingham et al 1993a; Dougados et al 1996*], and that there is a link between previous pain severity and continuous functional disability of the knee [*McAlindon et al 1993; Jordan et al 1997; Thomas et al 2008*]

2.3 Diagnosis of OA

The two main ways of identifying OA are symptomatic (clinical) and radiographic (x-ray) and they are as described below.

2.3.1 Symptomatic or clinical identification of OA

A joint can be diagnosed or identified with OA either by symptoms or structural changes or both [*Lawrence et al 2008*]. Generally, one does not need a special test to start treatment for OA in primary care – it is sufficient for a patient to present with symptoms such as pain, stiffness, tenderness, limitation of movement, crepitus (a crunching or grating sound or feeling) and/or occasional swelling of a joint [*Hordon et al 1993; Manek and Lane 2000; Murphy et al 2008*] in order to commence treatment. The symptoms of OA vary and can cause either short or long-term difficulty depending on the severity of the condition and the number of joints affected.

Knee OA usually causes pain while moving and walking, using stairs or rising from a bed or chair [*Altman et al 1986*] and hence leads to poor function of the knee. The symptoms tend to begin gradually and worsen over time. The pain is induced when the space between the bones narrows due to loss of cartilage and two bones press on each other.

OA of the hip may bring about pain in the groin, buttocks or thighs, which usually leads to limping or limited function of the hip [*Altman et al 1991*]. The pain induced by hip OA is sometimes referred to the knee and lower back, and as the disease progresses, movement of the affected hip can become limited. Hip OA can occur in either one or both hips [*Your Total Health 2008*]

Symptoms associated with OA of the hand include pain, swelling or enlargement of finger joints [*Altman et al 1990; Manek and Lane 2000*]. The joints in the hands that are typically affected are the base of the thumb and the two joints in the fingers [*AIHW: Department of Health and Ageing 2007*]. People with OA of the hand usually experience difficulty in coordinating the movements of their fingers (known as fine-motor movement) such as picking up items or gripping a pen [*Zhang et al 2002*].

Foot OA often causes pain and swelling in the foot particularly the joint at the base of the big toe. Symptoms may start with foot pain when wearing high heels or tight shoes, which previously did not cause any problems [*Your Total Health 2008*].

Most research has focused on regional pain (i.e. one joint site). However, many OA sufferers have generalized OA which is characterised by the involvement of two or more joints or groups of joints [*Gunther et al 1998*]. Generalized OA most commonly includes the knees, hips, fingers, and toes.

People with OA find it difficult and take longer time to perform daily activities (i.e. experience functional limitation), have less time available for leisure activities, depend much of the time on family and friends for assistance and spend more money on healthcare

than their peers of the same age and sex in the general population (*Yelin and Callahan 1995*). The type of activity a person with OA finds difficult to perform is greatly determined by which joints are affected. People with hand problems usually need help with self-care requirements such as household chores and dressing. When the hip or knee is affected, it leads to difficulty with bathing, dressing (especially undressing the lower part of the body), going up and down stairs, rising from a chair or bed, and walking [*Creamer et al 2000*].

Because pain and functional limitation are the main consequences of OA, they were recommended (in a consensus meeting with OA researchers and clinicians) to be used as the outcome measures for the prediction modelling study described in part 1 of this thesis. More detailed definition of these two outcome measures can be found in section 3.3.3 of chapter 3.

2.3.2 Radiographic identification of OA

The use of X-ray to examine structural changes in an OA joint is known as radiography. Such structural changes include the narrowing of the joint space, the thickening of the bones and/or the swelling of the joints [*Kellgren and Lawrence 1957; Croft et al 1990;*]. Most researchers' grade radiographs according to the Kellgren and Lawrence (K & L) scale [*Kellgren and Lawrence 1957*] which is used to classify OA according to the presence of osteophytes (i.e. outgrowth of bones usually at the margin of the knee and foot joints) and/or joint space narrowing.

Structural changes related to OA were observed on X-rays in more than 50% of persons over the age of 65, and almost unanimously in those age 85 years and over in North Sydney [*March and Bachmeier 1997*]. These structural changes are not always observed in people with joint symptoms and people with structural changes do not always have symptoms [*McDuffie et al 1987*].

The prediction modelling study in this thesis involves adults with symptomatic OA based on pain and functional limitations as they are the key symptoms for OA. The symptomatic diagnosis of OA is recommended in OA guidelines such as that provided by NICE and is generally used in primary care to make decisions regarding management of the consequences of OA.

2.4 Management of OA

Management of OA is aimed primarily at reducing pain, maintaining and improving the mobility and stability of affected joints in order to reduce functional limitations [*Altman et al 2000; Pendelton et al 2000; Manek and Lane 2000*]. The restorative options available to realise these goals are limited. The NICE guideline [2008] on the care and management of OA in adults recommends that three main interventions (education and advice, exercise and weight loss if overweight/obese) should initially be considered for people with OA. When further treatment is required, reasonably safe pharmaceutical therapies should be considered with surgical procedures considered as the last remedy.

Drug therapies used for OA management are simple analgesics (e.g. paracetamol), topical (e.g. capsaicin, diclofenac and rubefacients) or oral Non-Steroidal Anti-Inflammatory

Drugs (NSAIDs) (e.g. ibuprofen or COX-2 inhibitors) and local corticosteroid injections (e.g. methylprednisolone, hydrocortisone, triamcinolone). Successful management of OA can be achieved if a holistic patient-centred approach is applied during the assessment of the symptoms and signs (pain, stiffness, joint instability, etc) associated with OA and the psychological co-morbidities (anxiety and depression) that may influence these symptoms are also considered [*NICE 2008; NCC-CC 2008; Manek and Lane 2000*]. This can help to devise a high-quality management and control plan tailored to the needs of each individual suffering from OA. The above described approach to managing OA is similar to that proposed by both the European League against Rheumatism (EULAR) [*Combe et al 2007; Zhang et al 2005 and 2007*] and the American College of Rheumatology (ACR) [*Hochberg et al 1995a and 1995b*].

2.5. Prevalence of OA

Osteoarthritis is the most common form of arthritis in most countries and generally affects the knees, hips, hands and foot. Many people have degenerative changes in their joints, but due to absence of symptoms, this often goes undiagnosed. Population-based estimates of the prevalence of OA vary according to the definition used, the mean age of the population being investigated, the method used for evaluation, as well as the joint(s) involved. For example, estimates of prevalence rates vary depending on whether only moderate and severe symptomatic and/or radiographic changes are considered or if mild changes are also included [*Lawrence et al 2008*].

In recent decades, radiography has been used as the gold standard in the assessment of joints in most epidemiological studies [*Felson 1988; Peat et al 2001; Neame et al 2004*].

However, more recent clinical guidelines such as NICE [*NICE 2008*], EULAR [*Combe et al 2007; Zhang et al 2005 and 2007*] and ACR [*Hochberg et al 1995a and 1995b*] focus on clinical (symptomatic) OA as the most relevant definition for health care and for the patient.

The World Health Organization (WHO) estimate that 10% of the world's population over the age of 60 years suffers from OA and that 80% of the people with OA have movement limitations and 25% are unable to perform major daily activities [*WHO - Global Economic and Health Care Burden of Musculoskeletal Disease 2001a; WHO - World Health Report Archives 1995-2000. 2001*]. Most developed countries show increasing prevalence with age and higher prevalence among women than in men.

Table 2.1 summarises the prevalence of OA diagnosed either symptomatically or radiographically as reported in studies in the UK published from 1991 to 2005; most studies investigated prevalence of knee OA. The participants considered in the studies were aged 40 years and over with the majority of them being women except for the studies by Yoshimura et al [1998] and Birrel et al [2005], which included 13% and 48% women respectively. In summary, prevalence of OA irrespective of joint site ranged between 6.4% and 36.5% among the studies. Also, the prevalence rates based on symptomatic OA were generally higher than those based on radiographic definitions.

With regards to studies in other developed countries, estimates of OA prevalence again depended on the joints involved, population characteristics (age and gender distribution), and the method(s) used to diagnose it.

Table 2.2 below presents the findings of studies performed in other developed countries between 1990 and 2009. The participants were aged from 35 years and over and were mostly women. The methods of diagnosing OA employed in the studies were either clinical symptoms and/or radiographic measures mostly of the knee or hip joint. The prevalence of OA was generally higher among women than men except in the Chinese cohort examined by Nevitt et al [2002]. The prevalence of radiographic OA (using K & L ≥2) for both knee and hip joints ranges from 8% to 31% for men and from 6% to 47% for women. The highest prevalence of symptomatic OA for the knee and hip joints were reported in the Johnston County OA project in the USA, with estimates of 43% and 36% respectively.

In summary, the total prevalence of OA using K & L definition was slightly higher in the UK studies compared to the rest of the world. Similarly, the total prevalence of OA (based on only one study) using the Joint Space Width (JSW) definition was a little higher in the UK compared to the other developed countries with the exception of Denmark which used a different cut-off point (JSW<=2mm) instead of JSW <=1.5. However, total prevalence of OA was similar for both UK and the rest of the developed countries using symptomatic definition.

Table2.1 Prevalence of OA in the UK

Study	Definition of OA	No of participants	Age (years)	Gender (% female)	Prevalence (%)
Chingford, UK [Spector 1991]	Knee pain for most days in the past month	400	45–65	100	T=6.5
Calderdale, UK [Badley 1992]	Current knee joint problems (period unspecified)	15150	≥55	60	T=19
Bristol, Uk [McAlindon 1992a]	Knee pain for most days for at least one month	2102	≥55	60	M=20 F=27
Bristol, UK [McAlindon 1992b]	Knee pain + K & L ≥ 2	513	≥55	68	T=13
Nottingham, UK [O'Reilly 1996]	Knee pain in previous year on most days for at least one years	4057	40–79	52	T=25
Nottingham, UK [O'Reilly 1996]	Knee pain + K & L ≥ 2 + osteophytes	459	40–79	52	T=11
Manchester,UK [Urwin 1998]	Knee pain lasting at least a week in the previous month	3577	≥45	51	M=25 F=30
Manchester,UK [Urwin 1998]	Hip pain lasting at least a week in the previous month	3577	≥45	51	M=17 F=18
North Staffordshire & Southampton, UK [Yoshimura 1998]	Croft ≥ 3 (Hip)	1498	60–75	13	T=10.1 M=11 F=4.8
North Staffordshire, UK [Thomas 2004a]	Pain ≥ 1day in last 4 weeks (Hip), Pain ≥ 1day in last 4 wks (Knee)	7878	≥50	56	T=26.8, T=36.5
North Staffordshire, UK [Bedson 2005]	Knee pain and X-ray changes	146	≥45	66	T=12.5
South Manchester, UK [Birrell 2005]	JSW < 1.5mm (Hip)	1071	≥45	48	T=6.4

K & L – Kellgren and Lawrence; JSW – Joint Space Wide; T – Total; M – Male; F – Female

Table2.2 Prevalence of OA in other countries

Study	Definition of OA	No. of Participants	Age (years)	Gender (% female)	Prevalence (%)
Zoetermeer, Holland [Claessens 1990]	K & L ≥ 2 (Knee)	2865	≥45	54	T=20 M=15 F=24
Rotterdam, Holland [Odding 1998]	K & L ≥ 2 (Hip)	2895	≥55	60	M=14, F=16
	K & L ≥ 2 (Knee)				M=16, F=29
Arizona, USA [Hirsch 1998]	K & L ≥ 2 (Hip)	755	≥45	61	T=3.5
Iceland [Ingvarsson 2000]	K & L ≥ 2 (Hip)	3002	≥35	54	T=9
Beijing, China [Zhang 2001]	K & L ≥ 2 (Knee)	1787	≥65	59	M=28, F=47
Massachusetts, USA [Zhang 2001]	K & L ≥ 2 (Knee)	1,041	≥65	64	M=31, F=35
Beijing, China [Nevitt 2002]	JSW ≤ 1.5mm (Hip)	1506	≥60	60	T=1: M=1.1, F=0.9
Baltimore, Portland, USA [Nevitt 2002]	JSW ≤ 1.5mm (Hip)	316	60–74	51	T=4: M=5, F=4
Framingham Study Massachusetts, ISA [Zhang 2002]	Hand pain, aching or stiffness+K&L≥2	1,032	≥71	64	M=13, F=25
Dicomano, Italy [Mannoni 2003]	ACR criteria (Knee) ACR criteria (Hip)	697	≥65	58	M=30, F=8
Johnston County OA Project North Carolina, USA [Helmick 2008]	K & L ≥ 2 (Hip)	2637	≥45	57	T=27
Copenhagen, Denmark [Jacobsen 2004]	JSW ≤ 2.0mm (Hip)	3807	≥60	62	T=8
Rotterdam, Holland [Reijman 2004]	K & L ≥ 2 (Hip)	3585	≥55	58	T=7: M=8, F=6
Liège, Belgium, [Rabenda 2006]	Self-reported OA, GP contact for OA in last 6 months.	1811	Mean (SD) 51(6.6)	62	T=34
NHANES III, USA [Dillion 2006]	K & L ≥ 2 (Knee)	3128	≥60	53	T=3: M=31, F=42
Johnston County OA Project North Carolina, USA [Jordan 2007]	K & L ≥ 2 (Knee), Knee pain, aching or stiffness	3018	≥45	57	M=28 F=43
Johnston County OA Project North Carolina, USA [Murphy 2008]	Knee pain + K & L ≥ 2	3,068	≥45	57	T=5.4
Johnston County OA ProjectNorth Carolina, USA [Jordan 2009]	K & L ≥ 2 (Hip) Hip pain , aching or stiffness	3068	≥45	57	M=28, F=36

ARC - American College of Rheumatology; K & L – Kellgren and Lawrence; JSW – Joint Space Wide; T – Total; M – Male; F – Female.

2.6 Economic implications of OA

The frequency and chronicity of OA and the need for effective preventive measures make the disease a substantial economic burden for patients, health organizations, businesses, nations and the world as a whole [*Reginster 2002; Elders 2000; Lapsley et al 2001*]. It can result in a considerable loss of work time, as many people who suffer from the condition stay out of work either permanently or for a long period of time [*Leigh et al 2001*].

Most researchers reporting on economic consequences have included OA in the same category as other joint diseases or have grouped it with all other musculoskeletal conditions making it difficult to correctly estimate the economic cost of the disease [*Yelin 1998*]. In addition, different countries summarize their cost data using different monetary currencies, varied time period (monthly, annually etc) to which costs are allocated and varied calendar periods (with varied exchange rates or inflation rates) within which cost are calculated - which makes it difficult to compare the economic burden of OA between different countries. However, cost is usually converted into a common price year in order to overcome the problem of different currencies and the varied time periods of costing a condition.

The next two sections examine different types of OA related healthcare costs (direct and indirect costs) and the indirect costs related to loss of productivity.

2.6.1 Health care costs of OA

Direct health care costs refer to the costs that are associated with the provision of health care to OA patients. Examples of such costs include cost of labour (e.g. salary and benefits of health care providers such as doctors, nurses and physiotherapists), drugs, materials and equipment used for treatments, travel costs (for doctors and other health professionals), communication costs (telephone, fax, email, etc.) related to the provision of care to the patient and social care costs [*Drummond et al 2005; Tompa et al 2008*].

Indirect health care costs refer to informal care costs and other costs that cannot be easily linked to the care provided to the patient but are essential for the general operation of the organisation to facilitate efficient and effective performance of its activities. Such costs include overhead charges such as utilities (lighting and heating), fixtures and fittings, rent, administrative costs, rented equipment, etc [*Drummond et al 2005; Tompa et al 2008*].

Several studies have investigated direct or indirect health care costs associated with OA. The extent of the burden of OA in both developing and developed countries in terms of health care costs and lost wages are considerable [*Gabriel et al 1997a; Gabriel et al 1997b*]. In the years 2000 to 2001 there were over 2 million consultations with GPs concerning OA each year and over 114,500 hospital admissions for OA in the UK [*Arthritis Care 2004*]. The cost of health care under secondary and indirect care incurred by patients with OA was about twice compared to patients without OA [*MacLean et al 1998*].

In this chapter, all the costs of OA in other currencies were converted to UK 2010 cost and given in parentheses. The different currencies reported in this chapter were converted using the "Measuring Worth" website: http://www.measuringworth.com/ (Accessed 03-10 2010) and the "UK trade Information" website: https://www.uktradeinfo.com/index.cfm?Task=exchange&lastcountry (Accessed 03-10-2010).

In 2000/2001 the direct cost of OA was estimated at AU $1.2 billion (£566 million - converted to UK 2010 pound sterling), about 2.3% of the total allocated health expenditure of Australia for that financial year [*Access Economics 2001*]. The largest amount of this expenditure was attributed to hospital services (AU $567 million or £267 million), followed by residential old aged care services (AU $266 million or £125 million) and medications (AU $148 million or £69.8 million). Also, OA is the second leading reason for patient visits to a rheumatologist in Australia [*Lybrand 2003*].

Table 2.3 below summarizes the primary health care cost of OA reported in studies published from 1999 to 2009. Most of the participants sampled in the studies were women aged 50 years and over. The majority of the studies concerned knee OA with only one study involving participants with OA at any joint. The mean health care cost amongst participants with knee OA between 1996 and 2001 ranged from an average of £341 to £864 per person per year (estimated cost in 2010). Amongst those with OA at any joint, the cost was an average of £499/person/year, whilst the cost amongst patients with both knee and hip OA (£1,300) was almost double that of those who suffered OA at either their knee or hip joint (£871).

Table 2.3 Primary health care cost of OA

Country/ Study	Age (years)	Gender (% female)	Number of Participants	Joint affected	Description of care cost	Mean Cost /person /year for 2010
London, UK [Lord 1999]	62 (mean)	71%	174	Knee	Social direct cost (medication, primary healthcare services, transport and patient time.	[1997] £240 (£341)[a]
Indianapolis, USA [Mazzuca 1999]	63 (mean)	84%	94	Knee	Primary care visits and drug prescriptions for subjects receiving self-care education.	[1996] $1039 (£864)[a]
GISCA study, Italy [Leardini 2004]	66 (mean)	76%	254	Knee	Diagnostic procedures (visits and laboratory test), drugs, salaries and transportation.	[2001] €701 (£561)[ba]
Ontario, Canada [Gupta 2005]	≥55	74%	283	Hip and knee	Community services (e.g. transport, homecare, visiting nurse, meals) and paid help (e.g. for cleaning, shoveling snow, shopping)	[1996] CDN $2300 (£1300)[a]
Liège, Belgium, [Rabenda 2006]	51 (mean)	62%	617	Any joint	Visits to GPs, nurses, medical examination including radiographs and drugs including alternative therapy.	[2004] €534 (£499)[ba]
Artrocad study, Spain [Loza 2009]	≥50	74%	1071	Knee or hip	Medical, professional time, drugs and transport cost	[2007] €1116 (£871)[ba]

GISCA – Italian Group of Study of the Costs of Arthritis

2.6.2 Indirect cost of OA related to productivity losses

It was estimated that about 36 million working days were lost in 1999-2000 in Great Britain because of OA. This represents over £3 billion (£3.9 billion in 2010) in production lost [*Department for Work and Pensions 2002*]. Other investigators [*Yelin 1998; Praemer et al 1999; Elders 2000*] have calculated that OA cost more than $60 billion (£54.8 billion) per year in the US. This made OA second to ischemic heart disease as a cause of work disability in men over 50 years [*Lawrence et al 1998*]. The estimated cost of work-related OA ranged from $3.4 (£3.1) billion to $13.2 (£12) billion per year in the US (1994 dollars), making it as costly as work-related renal and neurological disease combined together [*Leigh et al 2001*]. Such estimates of economic consequences do not include pain and distress, adverse psychosocial effects, lost opportunities for increased productivity, inability to participate in regular exercise that could improve general health and reduce the cost to family members who cared for patients with OA [*Carr 1999*]. The economic cost for OA could be doubled if these unaccounted personal and social costs were included in the estimation of the cost of OA.

In Canada, Gupta et al [2005] reported that time lost from employment and leisure by subjects with disabling hip and/or knee OA aged 59 years and over and their unpaid caregivers accounted for about 81% of their total economic burden at CDN $12,200 (£6,950) per person per year.

In Belgium in 2004, an average of 0.8 sick days off work per OA patient each month was lost to productivity and this represents a mean cost of €64.5(£45.3) per OA patient per month [*Rabenda et al 2006*]. The total estimated cost of OA is about 1% of Gross

National Product (GNP) per year in France and most other developed countries [*Levy et al 1993; Doherty and Jones 1994*].

2.6.3 Summary of the economic implications of OA

As has been presented, the average estimated annual cost of OA per person varies among the joints involved as well as among countries. This cost represents a substantial part of health care resources and as such it must not be underrated. Although the prevalence of OA is higher at an advanced age, a significant number of economically active younger people may suffer with the condition [*Rabenda et al 2006*].

OA is a chronic condition and coupled with the fact that retirement age is increasing, it is plausible to expect that morbidity and economic impact of OA will increase with the apparent ageing of the world's population. This would result in most countries being faced with the dilemma of having a sizeable part of their labour force affected by OA. Thus the economic and social burden of OA particularly in the workplace is expected to increase [*Gupta et al 2005*].

Due to the large burden OA poses on the resources of healthcare systems and the economy of a country as a whole, this thesis explores the cost effectiveness of optimal primary care interventions for OA (described in chapters 8 and 9) to help decision makers make informed decisions in relation to the most cost effective option(s) to manage OA.

The prediction modelling study which is aimed at identifying high risk predictors of poor outcome of OA is presented in the next three chapters (3 to 5). Chapter 3 describes the background and methods employed for the prediction modelling study; chapter 4 describes the results of the prediction modeling study whilst chapter 5 presents the discussion and conclusion of the prediction modelling study.

Part 1

Chapter Three

Derivation of prediction models for OA – background and methods

This chapter outlines the objectives of the prediction modelling study as well as the study design and the target population used. The definitions of the two outcome measure used (severe pain and functional limitation) are provided and the various predictors used for developing the models presented. Also, an overview of the procedure generally used for deriving prediction models is described, the procedure followed to derive and internally validate the prediction models in this study is outlined and finally the criteria used to select the most important predictors of OA are presented.

3.1 Introduction

Predictive modelling involves identifying important predictors from a large set of candidate predictors (available at the time of prediction) which are supposed to be related to an outcome variable [*Shmueli and Koppius 2009*]. It generally leads to the development of dimension reduction models – i.e. the process aims to identify the combination of factors that best predicts an outcome of interest [*Hastie et al 2001, Moons et al 2009*]. The technique aims to develop models to provide accurate outcome prediction that can be applied in new prospective patient samples. It requires a systematic process during its development and validation in order to be applicable and efficient. The procedure generally followed to develop prediction models includes several steps including goal definition, data pre-processing, variable selection, model selection, model validation/performance evaluation and model reporting [*Shmueli and Koppius 2009*]. Detail description of the precise modelling choices (and justifications) employed in this study are presented later in this chapter.

Prediction modeling should be distinguished from an explanatory modelling framework, which is concerned with investigating the association between a specific predictor (or small number of predictors) and outcome based on hypotheses [*Hayden et al 2010*] and/or assumptions regarding the causal relationship with the outcome [*Shmueli and Koppius 2009*]. This approach aims to understand how and why certain medical conditions either occur or progress [*Gregor 2006*] - for example, investigating the association between obesity (adjusting for potential confounding factors) and progression of severe pain or functional limitation in adults with OA.

Predictive models can be valuable for medical practice and research purposes. In public health for instance, they may help to identify subjects at high risk of developing a disease who should be targeted for preventive interventions. A good example is the Framingham study of predictors for cardiovascular disease [*Wilson et al 1998*] which underpins one of the current policies for preventive interventions recommending that statin therapy should be considered only for patients with high risk of cardiovascular disease. In clinical practice, predictive models may inform physicians on the likelihood of a diagnosis or a prognostic outcome based on an individual patient's characteristics.

In this thesis focus is placed on the development of prognostic prediction models, identifying factors that strongly predict poor long-term outcome in people with joint pain and functional limitation. The resulting prognostic estimates may then be used to inform people with OA about the likely outcome of their condition, and to identify subgroups at high risk of poor outcome. Such information may help organizations such as NICE to develop policy guidelines on the care and management of OA in adults [*NICE 2008*].

Some predictors such as female sex, old age, obesity [*Zhang et al 2011; Urwin et al. 1998; Jinks et al 2008; Thomas et al 2008*] and baseline measures of pain [*Yusuf et al 2011; Dawson et al 2005*] have been shown to be strong predictors of the onset and progression of OA. Also, Thomas et al [2008] found that clinical history, physical examination and severity of radiographic knee OA add little value to generic factors in the attempt to predict whether older adults with knee pain will experience progressive or continuous functional disability which suggests that some factors are strong predictors whilst others are not.

Since no study has derived and validated prediction models to examine predictors of OA at several joint sites this study aimed to address these gaps in the literature following the aims described below.

3.2 Aims and objectives

The key aim of the predictive modelling study presented in this thesis is to identify prognostic factors associated with severe pain and functional limitation at three years follow-up.

The specific objectives pertaining to this aim are:

(i) To develop optimal models of OA - i.e. determine the optimal combination of factors associated with poor outcome of OA measured as severe pain and functional limitation after 3 years follow-up.

(ii) To examine the goodness-of-fit and performance of the models.

(iii) To internally validate the models.

(iv) To estimate population attributable risk (PAR) for each predictor – i.e. the maximum achievable health gain if optimal management would reduce/prevent the adverse effect of a predictor.

(v) To estimate number needed to treat (NNT) for each predictor – i.e. the number needed to treat to prevent one additional person from suffering with OA.

(vi) To use information on the strength of associations of predictors with outcome and PAR and NNT estimates to identify the most important set of predictors of poor outcome of OA.

3.3 Study population

In order to address the objectives of this part of the work, The North Staffordshire Osteoarthritis Projects (NorStOP) data was used. These are population-based (community dwelling adults) prospective cohort studies made up of participants aged 50 years and over registered with 8 general practices in North Staffordshire in the UK. A substantial proportion (98%) of the British population is registered with a general practice and hence this makes it a suitable sampling frame for a population study [*Bowling et al 1999*]. The main objective of the NorStOP was to study the clinical syndrome of OA in a general population sample of older people [*Thomas et al 2004b*].

The following subsections describe the design, composition and the recruitment process of this data, as well as how it will be utilized to meet the goals of this study.

3.3.1 Recruitment and sub-cohorts

The participants were selected through general practice records and recruited via a two-stage postal survey at baseline. The two-stage mailing strategy involves sending a Health Survey (HS) questionnaire to all subjects and then those who responded and gave consent to be contacted again and indicated that they had hand, hip, knee or foot pain in the last year were sent a Regional Pain Survey (RPS) questionnaire. For both surveys, those participants who did not respond after two weeks were sent a reminder postcard. In order to ascertain that the correct person had completed the questionnaire, the gender and date of birth given on the returned questionnaire were checked against data from the general practice list. Participants received additional questionnaires at 3 and 6 years [*Thomas et al 2004b*] using a similar recruitment procedure at baseline. Participants were recruited over three different periods of time leading to the formation of three sub-cohorts namely NorStOP1, NorStOP2 and NorStOP3.

The first cohort (NorStOP1) was recruited from three North Staffordshire general practices during April 2002 and consenting responders were followed up at 3 years (i.e. in April 2005). Responders who consented to the 3-year follow-up were followed up at 6-years (April 2008). The same procedure was followed to recruit participants for the NorStOP2 subcohort from July 2002 to August 2003 from three other general practices and the NorStOP3 from March 2004 to April 2005 from two other general practices. The NorStOP sub-cohorts 1, 2 and 3 were combined to carry out the analyses, using baseline and 3 year follow-up data only. Ethical approval was obtained separately for the recruitment of NorStOP1 and NorStOP2&3 projects and were all obtained from the North Staffordshire Local Research Ethics Committee (NS-LREC 1351 and 1430

respectively). The flowchart of the recruitment into NorStOP 1, 2 and 3 cohorts at baseline and three year follow up can be found in appendix 1a and 1b respectively.

3.3.2 Definition of the target population

The target population of this study was comprised of all respondents who reported any joint pain (hand, hip, knee or foot) at baseline for duration of 3 months or more in the past year with their 3 year pain and functional limitation scores used as their outcome measure. This definition of the target population was recommended in a consensus meeting with OA researchers, physiotherapists and GPs as an appropriate way of identifying people with symptoms of OA in the hand, hip, knee or foot. The rationale behind considering any joint site is because most people with OA have pain in multiple joint sites and focusing/treating only one site would potentially mean ignoring other pain problems when predicting the outcome of OA.

3.3.3 Definition of outcome measures

In the same consensus meeting during which the definition of the target population was recommended, outcome measures were also discussed.

As mentioned in chapter 2 section 2.3, pain and functional limitations are the two most common and most disabling consequences of OA and have also been recommended as two of the core outcome measures for OA [*Bellamy et al 1997*]. Hence, two main dichotomized outcome measures namely severe pain and physical function limitation (at three year follow-up) were adopted as suitable clinical endpoints for joint pain and OA and to identify subjects at high risk of poor outcome. In order to derive a single pain

score for each participant irrespective of the location of pain or the number of joints affected, scores from different pain questionnaires were used. Any participant with severe pain in at least one joint was classified as such.

The pain subscales of the WOMAC [*Bellamy 1996*], AUSCAN [*Bellamy et al 2002*] and FPDI [*Garrow et al 2000*] scores, scored by participants in the RPS questionnaire, were used to measure pain related to the hip/knee, hand and foot respectively. The WOMAC and AUSCAN are questionnaires which assess the three dimensions of pain, stiffness and physical function in hip/knee and hand using a sequence of 24 and 15 questions respectively. The FPDI instrument assesses four dimensions of pain, functional limitation, personal appearance and limitation in work or leisure activities in the foot using a set of 19 items. The WOMAC, AUSCAN and FPDI have been validated in several populations by Bellamy [1996], Bellamy et al [2002] and Garrow et al [2000]; Menz et al [2006]; Cook et al [2007]; Roddy et al [2009] respectively.

Limitations in physical function were assessed with the 10-item physical functioning subscale of the SF-36, which was completed by all participants (in both the HS and RPS questionnaire) regardless of the location of pain. The SF-36 contains 36 items [*Ware et al 1993*] which measures health on eight multi-item dimensions namely physical functioning (10 items), social functioning (2 items), role limitations - physical problems (4 items), role limitations – emotional problems (3 items), mental health (5 items), vitality (4 items), pain (2 items), and general health perceptions (6 items). The SF-36 questionnaire has been validated by Ware and Sherbourne [1992] and Brazier et al [1992].

Outcome measure for severe pain

To define a dichotomous measure of severe pain, cut-offs for high scores on the WOMAC, AUSCAN or FPDI scores at three year follow-up were selected using Receiver Operating Characteristic (ROC) curve analysis with a numerical rating scale (NRS) i.e. pain intensity score of 5 or higher as the anchor. The NRS score ranged from 0 to 10 and its validated cut-off point of 5 converts it into a binary variable as 0 – 4 (mild pain) and severe pain (5 – 10) [*Zelman et al 2003*]. .

Based on these ROC analyses, the cut-off points of the WOMAC pain score obtained for hip and knee pain were 6 and 5 (both ranges 0 to 20) respectively, 9 (range 0 to 20) for AUSCAN pain score and the FPDI pain cut-off was -0.479 (range -3.32 to 3.33). Participants with scores below the relevant cut-off point at three years follow-up were classified into the 'no or mild pain' category whilst those with scores equal to or above the relevant cut-off point were classified into the 'severe pain' category. Any participant with severe pain in at least one joint was classified as having severe pain whereas anyone who did not have severe pain in any of these joints was classified as having no or mild pain.

Outcome measure for functional limitation

The 10-item physical functioning subscale of the SF-36 (range 0 -100, higher scores indicating better physical functioning) was converted into a binary variable based on its median score of 55; median as opposed to the mean was used due to its distribution being heavily skewed. Respondents with a score of less than the median value were classified

as having severe functional limitation whilst those with scores greater than or equal to the median score were classified as having non-severe functional limitation.

3.3.4 Potential predictors

All the variables in the baseline HS questionnaire, including the baseline values of outcome variables (WOMAC, AUSCAN, FPDI and physical function), were considered as potential predictors in the predictive modelling study as most of them have been shown to be associated with symptoms of OA [*Thomas et al 2008, Jinks et al 2008*]. The baseline HS questionnaire included questions on socio-demographic factors, general health, joint pain in the last 12 months, participation restriction, social isolation, lifestyle characteristics, co-morbidity, medication use, and psychosocial factors. Full details of the variables and their composition have been described by Thomas et al [2004b], however a brief description is given below.

All the variables were treated as potential candidate predictors since every variable that contributes to a good outcome prediction is important. Variables with prevalence less than 10% or greater than 90% [*Tu et al 2001*] and those which were combined to create another variable (e.g. height and weight used to create BMI) were excluded from the analysis to facilitate successful convergence of the models and optimal discrimination between people with good or poor outcome. The variables with a prevalence lower than 10% included ethnic origin, spend most/all day in bed, play sports, take a bath/shower, heavy gardening, heavy DIY at home, access to telephone, access to chemist, access to bank, etc, and were excluded as they are not considered to be very useful when predicting outcome of OA and also will result in very low estimates of epidemiological indicators.

All continuous predictor variables were dichotomized, using predominantly the median as cut-off point due to the fact that the majority of the continuous variables were heavily skewed. The rationale behind doing this is that dichotomization often leads to simplicity of presentation of findings, easier understanding of results, and facilitates identification of subgroups of people at increased risk of poor outcome of OA. Co-linearity among variables was subsequently tested via chi-squared tests to help decide which variables to exclude from the analysis but the results generally showed weak correlation between the variables.

A total of 169 variables were identified at baseline in the HS questionnaire and the RPS questionnaire. Of these, 153 and 157 variables were included in the analyses of severe pain and functional limitation models respectively to identify the most useful set of predictors. The potential predictors used in the development of the models were classified under the following groups:

Socio-demographic factors

Binary socio-demographic factors were sex, social class, full time education, qualification obtained during adulthood and social class. Age, marital status and employment status were recorded as multi category factors. Body mass index (BMI) was also recorded as a categorical factor which was calculated as weight in kilograms divided by the square of height in metres, and obesity was based on BMI according to the World Health Organization (WHO) classification scheme - i.e. <20 kg/m^2 (under-weight), 20–24.9 kg/m^2 (normal weight), 25–29.5 kg/m^2 (over-weight) and > 30 kg/m^2 (obese) [*WHO 1997*].

Lifestyle characteristics

Lifestyle factors considered in this study were current smoking status with three levels (never smoked, previously smoked and currently smoking) and alcohol consumption with five levels (daily/most days, once/twice a week, once/twice a month, once/twice a year and never).

Severity and location of pain

Joint specific questions around the hand, hip, knee and foot were also included. For each of the above mentioned four joint sites, two questions within the HS on whether participants had problems and pain in last year were posed and were used to define binary variables indicating current pain in these locations.

Furthermore, the baseline WOMAC [*Bellamy 1996*] index scores for pain (range 0 to 20), stiffness (range 0 to 8), and physical function (range 0 to 68), and that for AUSCAN [*Bellamy et al 2002*] pain (range 0 to 20), stiffness (range 0 to 4), and physical function (range 0 to 36) were presented individually as binary variables using a cut-off point of below and above their median scores for easy interpretation with low scores being good and high scores being poor. Additionally, baseline FPDI [*Garrow et al* 2000] scores for pain (range -3.32 to 3.33) and function (range -4.05 to 4.32) were also presented individually as binary variables using their median value as the cut-off point with low being good and high being poor.

Bodily pain was measured using the pain manikin questionnaire made up of 50 items that covers the whole body. Participants were asked to shade on a blank manikin, any ache or pain that had been experienced for one day or longer over the last month. Each of the 50

items were classified binary variables and they are back (head, spine, lower torso, neck, lower), back left (shoulder, elbow, forearm, hand, upper torso, thigh, knee, calf, foot, hip), left (buttock), back right, (shoulder, elbow, forearm, hand, upper torso, thigh, knee, calf, foot, hip), right (buttock), front (head, throat, abdomen), front left (shoulder, elbow, forearm, hand, chest, thigh, knee, shin, foot, hip), front right **(**shoulder, elbow, forearm, hand, chest, thigh, knee, shin, foot, hip).

Use of pain medication

Questions on frequency of medication use (i.e. painkillers, creams, natural remedies and glucosamine /chondroitin) were originally 5 response items but were re-categorised into three groups namely 1=all/most days, 2=some days and 3=few/no, in order to achieve a better balance between numbers in each group.

Social isolation and participation restriction

Social isolation was measured with the Berkman-Syme Social Network Index (BSSNI) [*Berkman and Syme 1979*]. The BSSNI score ranges from 1 to 4 with 1 being the lowest score and implies least integrated and 4 being the highest score and implies most isolated.

Participation in social and other activities was assessed with the Keele Assessment of Participation (KAP) tool [*Wilkie et al 2002*] in the health questionnaire, designed to measure participation on eleven aspects of life including mobility within the home, such as self-care, and looking after other dependent. Questions were asked in terms of ability to perform different tasks in the past 4 weeks and the original 5 responses (all the time/most of the time/some of the time/a little of the time/none of the time) were

dichotomized into participation (a little of the time or none of the time) versus participation restriction. The resulting 11 binary items were then summed and dichotomized into 0 (no participation on any of the items) versus 1-11 (participation on 1 or more items)

Limitations in activity

Activity limitation was measured using 21 items within the health questionnaire on how often participants did certain activities during a normal day over the past 4 weeks, for example “go out for a walk”, “go to a club/church/social event”, etc. These items originally had 5 response categories but were collapsed into three responses (to have balanced numbers in the groups) as “all/most days in a week”, “some/few days in a week” and “no day in a week”.

Limitation in physical functioning and general health

The two components (physical and mental component scores) of the general health medical outcome survey short form (SF-12), validated by Ware et al [1996], both ranged from 0 to 100 (higher value indicating better score) and were dichotomized using their respective medians as cut off points.

Access to health care and social services

Measures of ability to access various types of services were included in the study as 7 binary variables in the health questionnaire, including for example “access to a car when personally needing it”, “good access to doctor as and when needed”, “access to advice or help with income”, etc.

Two further variables on participants' access to GP care were included in the study and were framed as "how often do you visit the doctor (GP) for yourself?" and "when you are ill, when do you go to your doctors?" These variables originally had 5 and 4 responses but were categorized into 3 responses for each question as often/very often/occasionally/seldom, hardly ever and straight away/wait a few days/wait several days/put off respectively.

Self-care and dependence on others

Six items on the health questionnaire assessed how the participants have changed the way they do their normal daily activities and whether they have depended on others to perform these, over the last 4 weeks. Example are "has the way they do things changed due to health compared to 12 months ago?", "assistance of others/aids required to move around the home", "assistance of others/aids required to go places outside of your home", etc. and were included in the study as binary variables.

Sleep

Participants' sleeping problems were assessed with validated questions [*Jenkins et al 1988*] concerning trouble falling asleep, waking several times per night, trouble staying asleep and waking up feeling tired over the past 4 weeks and were included in the analysis as three level (not at all/on some nights/on most nights) categorical variables.

Psychological factors

Psychological factors included symptoms of anxiety and depression measured with the 14 item Hospital Anxiety and Depression (HAD) Questionnaire [*Zigmond and Snaith 1983*], which have been validated with scores for both anxiety and depression ranging

from 0 (no distress) to 21 (high distress) – whose recommended NHS classifications are no HADS symptoms (0-7), mild HADS symptoms (8-10), moderate HAD symptoms (11-14) and severe HAD symptoms (15-21). However, in this study both symptoms were dichotomized as absence anxiety or depression (0-7) and presence of anxiety or depression (8-21).

Subscales of the Illness perceptions questionnaire (IPQ) [*Weinman et al 1996*] which include predictor attribution (score range 6 to 30), psychological attribution (score range 7 to 35), immunity attribution (score range 3 to 15), and accident/chance attribution (score range 2 to 10) were included in the analysis as binary variables as poor condition (below median scores) or good condition (above median scores).

The sickness impact profile (SIP) alertness scale, sometimes called cognitive complaints, was measured with the alertness score which ranges from 0 to 100 [*Bergner et al 1981*] and was also included in the analysis as binary variables as below (poor condition) or above (good condition) median scores.

Participants' views on health and life was assessed by asking participants how strongly they agree or disagree with 15 statements, for example "there is a lot which I can do to control my health", "what I do will affect whether my health gets better or worse", "treatments are effective in controlling disease", "my health is very unpredictable", etc. These originally had 5 responses but were re-grouped into three responses (to have balance numbers between groups) as strongly disagree/disagree, neither agree or disagree and agree/strongly agree.

Co-morbidity

A measure of comorbidity was obtained via health questionnaire item asking the participants whether they suffer with any of the following specific health problems: chest problems, heart problems, deafness, poor eyesight, raised blood pressure and diabetes. In addition, participants were asked whether they suffered with any of the following symptoms in the past 3 months: falls, difficulty remembering things, cough with spit, breathlessness when walking, dizziness/unsteadiness, weakness in arm/leg. All were modelled as binary variables.

3.4 Statistical methods

This section describes the application of generalized linear model (GLM) appropriate for modelling binary outcomes with emphasis on Poisson and logistic regression model techniques. Thereafter, the procedures employed to develop and test the prediction models in this study as well as the formulae used to calculate the epidemiological measures for each predictor identified by the models are illustrated.

3.4.1 Generalized linear models

GLM is the framework within which statistical techniques such as Poisson and logistic regressions are applied to model a non-normal outcome. GLM which was introduced by Nelder and Wedderburn in [1972], is a generalization of the classical linear model which enables a model to be fitted to data that do not follow normal distribution - a frequent

occurrence in medical research where one often encounters binary, ordinal, count, survival and many other types of non-normal data. The generalization also relaxes the assumption of homogeneity of variances, an assumption required by the standard linear regression.

GLM takes the following general mathematical form:

$$E(Y) = g(\mu) = \beta_0 + \beta_1 x_1 + ... + \beta_p x_p \qquad (1)$$

where **g(μ)** is the link function relating the linear component on the right hand side to the outcome variable; β_o is the intercept; $\beta_{1,...,}\beta_p$ are regression coefficients measuring the effect of variables $x_{1,...,}x_p$ respectively.

Table 3.1 gives link functions for commonly encountered distributions of data.

Table3.1 Common link functions

Outcome	Probability distribution	Link function
Continuous	Normal	Identity
Count	Poisson	Log
Binary (0/1)	Binomial	Logit
Ordinal	Multinomial	Log

Estimation in the GLMs is performed via maximum likelihood estimation (MLE). In order to briefly outline the procedure, assume there are *n* independent identically distributed observations in a study taken from some probability distribution ***f*(.);** then the likelihood function of parameters of interest ($\theta = \beta_{0,}\beta_{1,...,}\beta_p$) is given by:

$$L(\theta \mid x_1,...,x_n) = \prod_{i=1}^{n} f(x_i \mid \theta) \qquad (2)$$

It is however generally more convenient to work with the natural logarithm of the likelihood function, given by:

$$\ln L(\theta \mid x_1,...,x_n) = \sum_{i=1}^{n} \ln f(x_i \mid \theta) \qquad (3)$$

The value of θ that maximizes the likelihood (and log-likelihood) is called the maximum likelihood estimate. MLE is a numerically intensive iterative procedure, implemented in STATA (the software used in this thesis) via Newton-Raphson procedure. The process begins with specification of initial parameter values; usually the null value, and then a quadratic approximation to the likelihood function around these initial values are constructed. Subsequently, adjustments are made to parameter values so that this quadratic approximation is maximized. This iterative process stops when the parameter values have stabilised, and it is said that convergence has been reached. Most computer programs record the value of likelihood function at each iteration performed to achieve convergence. Sometimes the program fails to achieve convergence and this is usually due to insufficient data to support the estimation of the parameters in the model. Non or slow convergence may also be due to other issues such as co-linearity between variables, but in STATA this is not a problem because the program automatically detects co-linearity before the iterative process begins and discards any such problematic variables.

In this study, Poisson regression model was selected as the primary model of choice because it estimates the true risk i.e. relative risk (RR) for each predictor which was subsequently used to calculate population attributable risk (PAR) and number-needed-treat (NNT) for each predictor. The number of events in this study was relatively high and as such odd ratios (ORs) from logistic regression analysis would not approximate

relative risk (RR) very well in this situation [*Zhang and Yu 1998*]. However, the logistic model was also employed for comparative purposes. Both Poisson and logistic models are now described in more detail.

The conventional p-value of <0.05 was chosen as the significant level for retaining variables in all the analyses carried out in this thesis. All the analyses were performed using STATA version 11 [*StataCorp. 2009*].

3.4.1.1 Poisson regression model

Poisson regression analysis is a popular model for count data. One of its original applications was to model the number of deaths of Prussian soldiers from horse accidents [*Bortkewitsch 1898*]. It has also been applied extensively in routine laboratory work which includes the monitoring of radioactive tracers by emission counts and the count of infective organisms observed on a slide under a microscope [*McCullagh and Nelder 1989*]. In medicine it has been used to model the incidence of diseases such as OA [*Dominick et al 2005*] and how such incidence rates may be affected by factors such as age, gender, social class, and other exposures. Its distribution describes the number of occurrences of an event over a specified period of time or region of space provided that the events occur independently and at random [*McCullagh and Nelder 1989*]. The assumptions of Poisson regression are that:

1. The exposure variable(s) are linearly related to the log rate of the expected value of outcome (i.e. logarithmic link function): E.g. $\log(\mu) = \beta_0+\beta_1x_1+\ldots+\beta_px_p$.

2. The expected value of the outcome is a multiplicative function of the exposure variable(s). E.g. from point one above; $\mu= \exp(\beta_0+ \beta_1x_1+ \beta_2x_2) = e^{\beta}e^{\beta 1x1}e^{\beta 2x2}$.

3. The expected count and the variance are equal.

4. The observations are independent.

Note that RRs are obtained by taking the exponential of the estimates of the β's (i.e. regression coefficients)

Robust variance estimator was used to determine uncertainty of the effect estimates [*Zou 2004*] as it ensures that small deviations from the model's standard assumptions (such as non- linear relationship between outcome and exposure variables, over/under dispersion of data) are controlled for and hence provides more accurate effect estimates and their precision.

3.4.1.2 Logistic regression model

The logistic model is used for modelling binary response. It was first proposed for use in demographic studies by Verhulst (1838) and given its present name by Reed and Berkson (1929). It has been used extensively in the fields of marketing applications for predicting a customer's propensity to purchase a product or stop a subscription [*Agresti 1996 and 2007*]. It has also been used in medicine and social science to examine predictors for onset or progression of conditions such as OA [*Jinks et al 2008; Wilkie et al 2007; Peat et al 2004; McAlindon et al 1993*].

In terms of the GLM formulation outlined, for logistic regression, the logit link function is used to linearly relate predictor variable(s) to the log odds of the outcome:

$$E(Y) = logit(\mu) = \beta_0 + \beta_1 x_1 + ... + \beta_p x_p \qquad (4)$$

where logit (μ) = (μ/1-μ) and μ is the proportion of subjects with the outcome.

ORs are subsequently obtained by taking the exponential of the log odds values of each predictor.

In order to calculate the PAR and NNT (both dependent on the RR for their calculation) for each predictor in the logistic regression models, the formula used by Zhang and Yu [1998] to convert ORs into RRs was adopted:

$$RR = \frac{OR}{(1 - P_0) + (P_0 \times OR)} \qquad (5)$$

where $\boldsymbol{P_0}$ indicates the proportion of the outcome in the unexposed group.

3.4.2 Variable selection procedure

Different methods of variable selection have been proposed that can be applied in prediction modelling. However, they do not produce the same results when applied to the same problem and there seem to be no consensus as to the best overall approach since all the methods may be criticized for different reasons. In practice, either a full model (all variables) is fitted or one out of the several variable selection methods is used. There are two main types of variable selection methods namely stepwise and all subset methods.

The stepwise methods include forward and backward selection and usually adopt a selection criteria of $p<0.05$ in advance to sequentially enter or remove variables from a model. For forward selection method, the model starts with no variables and then at each

subsequent step the variable whose contribution leads to the greatest improvement (significant at $p<0.05$) in the fit of the model is added until all included variables demonstrate significant improvement in the model with p-value less than 0.05. On the other hand, backward selection begins with the full model (i.e. includes all variables) where at each step the variable with the lowest contribution ($p>0.05$) to the fit of the model is excluded. The process terminates when all the variables remaining in the model significantly improve the model with p-values < 0.05.

The combination of the forward and backward selection strategies is also possible but applicable only in a few advanced statistical software such as STATA and SAS. This technique starts with forward selection then after the inclusion of the second variable; it tests at each step whether a variable already included can be removed from the model with a significant decrease in the model fit.

The second method, the all possible subset method, is an extension of the stepwise selection method. This is where all (2^k - where k is the number of variables) possible subsets (combination of predictors) of models are derived for a given data and the best model is subsequently chosen by using information optimization criteria from the likelihood function such as the Akaike information criterion (AIC) or Bayesian information criterion (BIC). Adopting this approach in this study however was not practical due to the large number of potential predictor variables that were considered.

Because the backward selection procedure has been found to be efficient in selecting a more stable combination of predictors for a model [*Steyerberg 2009*] compared to the forward selection methods, it was used to construct the models in this study using a

standard significance level of $p<0.05$ for retaining variables in the models based on the likelihood ratio (LR) test.

3.4.3 Optimum model selection

The use of an appropriate variable selection technique helps to identify the right number of predictors for a model which ultimately helps to accurately estimate the predictive performance of the model [*Hastie et al. 2001*]. In general, information criteria methods such as AIC and BIC are used to compare and select the best model when there are several competing models to choose from and the model with the smallest AIC or BIC value is preferred. These information criteria methods are based on the goodness-of-fit of a model and are penalized by the complexity (i.e. number of predictors) of the model.

The AIC is defined as [*Akaike 1974*];

$$AIC = -2LL + 2p \tag{6}$$

where;
LL is the log-likelihood value of the derived model
p is the number of estimated parameters

The BIC is also defined as [*Schwarz 1978*];

$$BIC = -2LL + p\log n \tag{7}$$

where,
LL and *p* are as above and *n* is the total number of observations.

Both AIC and BIC measures were used to compare the Poisson and logistic regression models for the pain and functional limitation outcomes in this study.

3.4.3.1 Missing data and multiple imputation (MI)

Missing data can occur in any type of health research and when they are not dealt with appropriately can lead to biased estimate of the association being examined [*Little 1992; Greenland and Finkle 1995*]. Missing data are often unavoidable especially when one is dealing with large cohorts such as the NorStOP cohort which measured many predictors. When the missingness of data does not depend on any observed data (i.e. outcome or other predictors) in the dataset it is known as missing completely at random (MCAR). An example of data MCAR is when the data collected is accidentally lost or destroyed (e.g. page containing a question is soiled or ripped off). When the data missing is dependent on some observed predictor variable but not on the outcome variable of interest it is known as missing at random (MAR) [*Schafer* 1997]. For example, men are likely to miss a question linked to pregnancy since such a question can only be answered by women as it is dependent on one's gender. Data not missing at random (NMAR) usually occurs when the data missing depends on the actual value of the missing data [*Rubin 1976*]. For example, when a subject is asked for his/her income level it likely that they might refuse to answer such a question if they have relatively high income.

The techniques used for dealing with missing data range from simple techniques such as complete case analysis (ignoring missing values), using preceding value in the dataset (i.e. last data carried forward), imputing mode value, overall mean imputation, single

regression (i.e. the missing value used as outcome variable and all other data points for an individual used as predictor variables) to complex techniques such as multiple imputation (MI). In general the simple techniques are inadequate as they lead to biased estimates whilst the more complex method produce reasonably better results by minimising this bias [*Rubin 1987; Greenland and Finkle 1995*]. When each entry of missing data is imputed only once, it is known as single imputation whereas MI is where several or multiple imputed data sets are created for each missing data to estimate its value using chain regression equations [*Rubin 1987; Schafer 1997*].

The multiple imputation by chained equation (MICE) technique uses an algorithm where the first variable say x_1 with at least one missing value is regressed on other variables say $x_2,\ldots,x_k$ with the estimation restricted to subjects with valid values for x_1. The missing values in variable x_1 are then replaced by simulated random draws from its underlying distribution given the other variables. This process is repeated for all the other variables with missing data in turn and is most commonly repeated 5-10 times to produce such many imputed datasets in which standard statistical techniques are then applied to produce effect estimates with standard errors and then combined appropriately. MI generates data that may be costly to collect and in a form in which standard statistical methods could be used to analyse.

In this study, about 47% of the baseline variables used (n=75/159) had a degree of missing values, ranging from 0%-18% of observations (average=3%, mode=2%). The primary analysis was based on a complete-case analysis, but MICE technique was carried out to generate 10 imputed data per missing value for nine baseline variables with missing values greater than or equal to 3% (i.e. the average proportion of missing

values). The results (composition and performance of the prediction models) based on the imputed datasets were compared to the results of the complete case analysis.

3.4.4 Goodness of fit

Goodness-of-fit is sometimes referred to as calibration and examines how well a model's observed values agree with its expected (predicted) values. This is an indication of how well a statistical model fits given data. Several methods can be used to assess the goodness-of-fit of a model which can be classified into Chi-square analysis methods (i.e. Pearson goodness of fit test and Hosmer–Lemeshow goodness-of-fit test) and Deviance analysis methods (i.e. calibration slope and shrinkage).

Pearson goodness-of-fit test tests the null hypothesis that the frequency distribution of the events observed in a sample is consistent with the expected distribution [*Chernoff, and Lehmann 1954*]. It is calculated (using the chi-square statistic) as the squared difference between each observed and expected frequency for each possible outcome, divided by the expected frequency and summed over all outcomes. The statistic (chi-square) can be used to calculate a p-value by comparing the estimate of the statistic to a chi-square distribution under the appropriate degrees of freedom. P-values greater than 0.05 imply good fit and vice versa. The number of degrees of freedom is calculated as number of rows (e.g. levels in a categorical predictor variable) minus 1 multiplied by number of columns (e.g. levels in outcome variable) minus 1.

Hosmer–Lemeshow goodness-of-fit test examines if observed event rate matches expected event rate in subgroups of the model sample. Usually, deciles of the data are used as the subgroups [*Hosmer and Lemeshow 2000*]. This test produces a p-value which determines the model fit – a higher p-value ($p > 0.05$) indicates a better fit whilst $p < 0.05$ indicates lack of fit. This method is not as efficient as the Pearson goodness-of-fit test because it uses subgroups to examine the fit of a model [*Harrell et al 1996*].

Calibration slope is the graphical representation of the agreement between observed values (on the y-axis) and expected values (on the x-axis) of a regression model with perfect agreement being the 45-degree line (i.e. with an intercept 0 and slope 1). In the case of linear regression models the calibration slope is a scatter plot between the observed and expected values whilst for models with binary outcomes (assuming y=0 or y=1) smoothing techniques such as lowess are used to estimate the expected values of the outcome (y=1) by averaging response values with similar expected values.

In this study, goodness-of-fit were assessed by examining the calibration plots of the models as well as their estimates of Pearson Goodness-of-fit test. These techniques were used as they are most commonly applied for generalized linear models.

3.4.5 Model validation

This section describes the process and techniques generally used to validate a prediction model.

Model validation is the process by which the predictive performance of a model is determined, in order to estimate accuracy of model predictions when used in future, prospective samples. It is an important step in the process of model development as it estimates the ability of the model to identify subjects at increased risk of an outcome and also helps to check a given model for over-fitting (i.e. unstable effect estimates).

There are two main types of validation namely internal and external validation and they are described in the next two sections below.

3.4.5.1 Internal validation

Internal validation is the methods by which the prediction accuracy of a model is tested in the same sample used to derive the model. In general, the performance of a prediction model is overestimated when determined on the sample used to derive the model [*Steyerberg et al 2001*].

Various internal validation techniques are available that are aimed to provide a more accurate estimate of model performance in new rather than the same patients. The techniques are split-sample, cross-validation and bootstrapping.

Split sample validation involves randomly splitting the sample into two groups - one group being used to develop the model and the other group used to measure its performance. Typically, splits of 50% : 50% or 2/3 : 1/3 are used [*Steyerberg 2009*]. With this technique, model performance is determined on similar but independent data. The drawbacks to this technique are twofold: firstly, as samples are split at random, major variations may occur regarding the distribution of predictors and outcome as only

one part of the data, as opposed to the entire data, is used for model development [*Molinaro et al 2005*]. Secondly, as the validation data is relatively small it may lead to unreliable assessment of model performance. That is, a partial performance estimate is obtained given that only a part of the data is used even though one wants to know the peformance of a model based on the full sample.

An extension of the split sample method is cross validation technique, where the model is once again developed on one randomly drawn half of a sample and tested on the other and vice versa. This process is usually repeated several times and the average of the test estimates is taken as model performance, thus producing a more reliable model performance estimate compared to split sample validation.

Bootstrapping technique was originally developed by Bradley Efron [1979; 1981; 1982] and further developed by Efron and Tibshirani [1993a]. It is a computer intensive re-sampling technique in which large numbers of samples are drawn with replacement from an original sample, with each drawn sample being of the same size as the original sample. The technique is usually used when the distribution of a statistic is complicated or unknown and also when the sample size is not large enough to perform a standard statistical inference.

Because bootstrapping provides relatively unbiased estimates of performance estimates (prediction accuracy), and uses the whole dataset for model development it was used as the internal validation tool in this study [*Steyerberg et al 2003*]. The technique was only used to validate the models.

In order to estimate optimism, the procedures used to develop the prediction models were replicated in each of the 500 bootstrap samples and estimated the performance in each sample [*Steyerberg et al 2003*]. These were adjusted in the original sample to estimate test performance where the difference between the bootstrap and test performances is the optimism value.

The Poisson and Binomial families as well as the log and logit links were specified for the respective Poisson and logistic regression models.

3.4.5.1.1 Performance evaluation

Several approaches have been proposed to examine the performance accuracy of a model. These include techniques such as C-statistics (discrimination method), R-squared, Shrinkage and Brier score.

Discrimination is the ability to correctly classify subjects with or without an event of interest. The most common measure of discrimination is the Receiver Operating Characteristic (ROC) curve which is a plot of sensitivity (true positive rate) on the y-axis versus 1-specificity (false positive rate) on the x-axis. The area under the ROC is identical to the C-statistic (C-index) for a binary outcome and ranges from 0.5 (no predictive ability or discrimination above chance) to 1 (perfect predictive ability or discrimination). C-statistics estimate < 0.7 is considered as moderate discrimination whilst an estimate ≥ 0.7 is considered as good discrimination [*Hosmer and Lemeshow 2000*].

The R^2 is the most common performance measure for continuous outcomes. It represents the proportion of variation in an outcome that is explained by the predictors in the model. For generalized linear models, the Nagelkerke's R^2 is commonly used [*Nagelkerke 1991*].

Shrinkage is the pulling together (flattening) of the plot between observed and expected values to the 45-degree line and it occurs as a result of over-fitting. It is that constant value used to multiply the regression coefficients (excluding the intercept) needed to make a model perfectly calibrated for future samples. The heuristic estimator of shrinkage [*Van Houwelingen and le Cessie 1990*] is defined as λ = Model χ^2–*p* / Model χ^2; where *p* is the number of regression coefficients (excluding intercept but including interaction effects) and Model χ^2 is the likelihood ratio chi-square statistic of the fitted model (i.e. -2 times the difference in log-likelihood between the null model and fitted model). A shrinkage value of 1 implies no over-fitting or absence of over-optimism. The value of shrinkage decreases when large numbers of predictors are considered and vice versa. For linear regression models, shrinkage is estimated as the ratio of the adjusted R^2 to the ordinary R^2.

The Brier score [*Brier 1950*] can also be used when the outcome of interest is binary. It is the average squared difference between the observed outcome and the expected outcome (i.e. $(\sum O_i - E_i)^2 / n$ - where O and E are the observed and expected outcomes for each subject and *n* is the number of subjects). This score can range from 0 (for perfect model) and 0.25 (for uninformative model) when the incidence of outcome is 50% or more

whereas when the incidence of outcome is 10% the uninformative estimate reduces to 0.09 [*Steyerberg 2009*].

The estimate of a performance measure such as the c-statistic is more optimally determined by an appropriate internal validation technique such as bootstrapping. In order to examine the optimism of a model, (i.e. if a model is over or under fitted) the fitted model's performance estimate is compared with the bootstrap model's performance estimate. A positive difference for example, between the c-statistic values of the fitted and bootstrap models indicates over-fitting (over-optimism) whilst a negative difference implies under-fitting (under-optimism). When these occur it implies that a model is unlikely to be valid in future subjects from a similar population [*Harrell et al 1996; Steyerberg 2009*]. The performance estimate of a derived model is usually higher than that in future patients' model and this problem is particularly common in small datasets with relatively few subjects with the outcome or few outcomes compared to the number of variables [*Harrell et al 1984; Laupacis et al 1997*].

In this study c-statistic was used as the performance measure for the prediction models as it produces reliable estimates compared to the other measures and also is commonly used for binary outcome models. Optimism was examined in both the Poisson and logistic regression models and were calculated based on the components stated below [*Steyerberg et al 2003; Steyerberg 2009*];

(i) Apparent performance: refers to the estimate of a model performance from the entire sample (final, best model selected).

(ii) Bootstrap performance: is the average estimate of model performance based on prediction models derived from the 500 bootstrap samples. The same procedure used to derive the prediction models were used in each of the bootstrap samples.

(iii) Test performance: is the average estimate of model performance after applying the regression coefficients from the prediction models derived from the 500 bootstrap models to the original (entire) sample.

(iv) Expected optimism (EO): is the difference between the estimated bootstrap performance and test performance.

(v) Optimism corrected performance (OCP): is the difference between the estimated apparent performance and the optimism.

3.4.5.2 External validation

This is where the prediction accuracy of a model is tested in a sample different from that used to derive the model with the aim of generalizing the results to new patients. It is considered as the gold standard for model validation because in general, a model performs better on the data used for its development compared to the performance of the same model on new data [*Harrel et al 1996; Efron 1983*]. As a result external validation is performed on datasets of subjects that are different in some respect (e.g. location, clinical setting, etc) compared to the data used to derive the model. Although internal validity can give some indication of the optimism of a prediction model, it may not provide true estimate of the performance of the model when implemented in other settings or populations. Due to lack of access to external data that includes similar information on predictors as well as the outcome measures used in this study, it was not

possible to carry out external validations for this thesis but it would be considered in future work.

3.4.6 Calculation of maximum health gains: (PAR) and NNT

The aim of this section is to explain how population attributable risk (PAR) and number needed to treat (NNT) were calculated and how these were used to identify the most important predictors of poor outcome of OA. PAR and NNT were calculated for each predictor. Details of the definitions and calculations of these two epidemiological measures are given below.

3.4.6.1 Population attributable risk (PAR)

This is the proportion of the incidence of an outcome (e.g. severe pain) in the total population that is due to exposure to a specific predictor. It can also be interpreted as the proportion of the incidence of the outcome in the whole population that would be eliminated if the exposure to such a predictor was completely eliminated (e.g. by successful intervention).

The unadjusted PAR is calculated as:

$$PAR = \left(\frac{\pi - \pi_0}{\pi} \right) \qquad (8)$$

where

π is the proportion of the outcome among both exposed and unexposed subjects

π_0 is the proportion of the outcome among the unexposed subjects

The corresponding 95% CI for unadjusted PAR is calculated using the formula derived by Lui [2004] based on the delta method, a method generally used for calculating the variance of complex functions. A summary of Lui's derivation taken from Hildebrandt *et al* [2006] is given below.

The PAR formula given in equation (8) can be expressed as;

$$PAR = \left(\frac{\pi - \pi_0}{\pi} \right) = 1 - \Theta \tag{9}$$

where,

$$\Theta = \left(\frac{\pi_0}{\pi_{0.}\pi} \right) \tag{10}$$

where,

π_0 and π are as described above (in equation 8) and

$\pi_{0.}$ is the proportion of unexposed subjects

Using the delta method the asymptotic variance for Θ is;

$$VAR(\Theta) = \Theta^2 VAR\left(\log(\Theta)\right) \tag{11}$$

with,

$$VAR\left(\log(\Theta)\right) = \frac{1-\pi_0}{N\pi_0} - \frac{\pi_{0.} + \pi - 2\pi_0}{N\pi_{0.}\pi} \tag{12}$$

where π_0, $\pi_{0.}$ and π are as described above and N is the total number of subjects.

Thus, the 95% CI for PAR directly based on Θ is given by;

$$\left[\left\{1-\Theta-z_{1-\frac{\alpha}{2}}\sqrt{VAR(\Theta)}\right\},\min\left\{1-\Theta+z_{1-\frac{\alpha}{2}}\sqrt{VAR(\Theta)},1\right\}\right] \quad (13)$$

hence

$$95\%CI(PAR)=\left[UL(PAR),LL(PAR)\right] \quad (14)$$

where *UL*(*PAR*) and *LL*(*PAR*) are the upper and lower confidence limits of the population attributable risk.

STATA command aflogit was used to estimate adjusted PAR with 95% CI. The command uses the generalized formula given by Greenland and Drescher [1993] for estimating adjusted PAR for a categorical predictor denoted by;

$$PAR=1-\pi_{1j}\left(\frac{1}{RR_j}\right) \quad (15)$$

where

π_{1j} is the proportion of the outcome among exposed subjects in stratum j

RR_j is the relative risk comparing stratum j with stratum 0 (i.e. the reference group)

The variance of the above measure has no simple form and aflogit uses the delta method used by Greenland and Drescher [1993] which is computationally intensive to carry out.

Predictors with high PAR indicate factors associated with a high risk of poor outcome, and take into account the proportion of the population exposed to this factor. PAR is therefore a useful measure to select relevant predictors in a given population, combining

information on the occurrence of the risk factor in the population and strength of the association with outcome.

3.4.6.2 Number needed to treat (NNT)

NNT represent the number of people needed to be treated to prevent one additional person from suffering with a poor outcome of OA.

The unadjusted NNT formula proposed by Heller et al [2002] was used:

$$NNT = \frac{1}{ARI} = \frac{1}{\pi_1 - \pi_0} \qquad (16)$$

where

ARI is the absolute risk increase, measuring the increase in risk of an outcome in an exposed group compared to an unexposed group.

π_1 is proportion of a outcome among the exposed subjects.

π_0 is the proportion of outcome among the unexposed subjects.

The relevant formulae for calculating 95%CI for ARI are given by Heller et al [2002] and are reproduced here as given by Lui [2004] and Hildebrandt et al [2006]:

$$VAR(ARI) = \frac{\pi_1(1-\pi_1)}{N_1} + \frac{\pi_0(1-\pi_0)}{N_0} \qquad (17)$$

Where N_0 is the number of unexposed subjects and N_1 is the number of exposed subjects.

The 95% confidence interval for *ARI* is given by:

$$\left[\max\left\{ARI - z_{1-\frac{\alpha}{2}}\sqrt{VAR(ARI)}, -1\right\}, \min\left\{ARI + z_{1-\frac{\alpha}{2}}\sqrt{VAR(ARI)}, 1\right\}\right] \qquad (18)$$

Hence the 95% CI for the NNT is given by:

$$95\%CI(NNT) = \left(\frac{1}{UL(ARI)} to \frac{1}{LL(ARI)}\right) \qquad (19)$$

Where *UL*(*ARI*) and *LL*(*ARI*) are the upper and lower confidence limits of the absolute risk increase.

The delta method would have been used to compute adjusted NNTs with 95% CI but because this is too computationally intensive, adjusted NNTs were not calculated.

Assuming that the 95% CI of an NNT estimate reports a negative upper limit, it means that it is not statistically significant. Denoting the NNT(95%CI) for predictor X as;

X = 20(10 to -15)

This can be re-written as;

X = 20 (NNB 10 to ∞ to NNH 15)

where,

NNB – Number needed to benefit

NNH – Number needed to harm

This 95% CI can therefore be interpreted as ranging from a NNT for one person to benefit (NNB) from 10 to infinity, to the number treated for one person to be harmed being 15.

3.4.7 Identification of most important predictors

The rules used to identify important predictors for a particular condition vary and include selection of high effect estimates (e.g. ORs and RRs), high epidemiological indicators (e.g. PAR) and high predictive probability of a predictor.

In this study the most relevant predictors for severe pain and poor functional limitation at three years were selected using the selection rule employed by Smit et al [2006] to select subgroups of participants at increased risk of developing anxiety in later life. The rule selects predictors based on large effect size (IRR), high PAR and low NNT which has the ability to select factors for which the highest possible health benefit (IRR and PAR) and the lowest possible effort and cost (NNT) can be achieved if interventions are completely successful. In addition, the performance estimate of the set of predictors that would be chosen based on these criteria must be comparable to that of the respective original Poisson and logistic regression models. The identified predictors may be used in future for easy identification of high risk subgroups that are at high risk for poor long-term outcomes of OA.

Chapter Four

Derivation of prediction models for OA - Results

This chapter outlines the main findings of the prediction modelling study. It includes a description of baseline characteristics of the participants in the NorStOP cohorts as well as an illustration of the development of Poisson and logistic regression models for severe pain and functional limitation at three years. The composition and performance of the models developed are presented in a logical order to reflect the procedure outlined in the previous prediction modelling methods chapter.

4.1 Response rates

A total of 26705 people were identified from the eight general practices aged 50 years and over, out of which 26625 were eligible to take part in the NorStOP cohort and were mailed the health survey (HS) – 18497 (71%) of these responded. After excluding refusals and non-consents, a regional pain survey (RPS) was sent to 10057 participants, of whom 8734 (87%) responded.

After the baseline stage of the study, 12641 people were eligible for the 3-year follow-up. After excluding relocated and dead participants, 11918 were mailed the HS questionnaires of which 9705 (adjusted response rate 81%) responded. RPS questionnaires were mailed to 6611 (excluding no consent/dead/refused) participants, with 6181 questionnaires (adjusted response rate 93%) completed and returned.

Full details of recruitment into the NorStOP cohorts at baseline and 3 years are shown in the flowcharts, figures 4.1a and 4.1b respectively.

The results of analyses presented in this chapter are based on the responders to both the baseline HS and RPS as well as the follow up HS and RPS at 3-years which included 6181 participants. Of these, 3563(57.6%) satisfied the criteria for the target population (see chapter three section 3.3.2) by reporting chronic pain at baseline which lasted for a period of 3 months or more in the previous year.

Figure 4.1a. Flowchart of recruitment into the NorStOP cohorts at baseline.

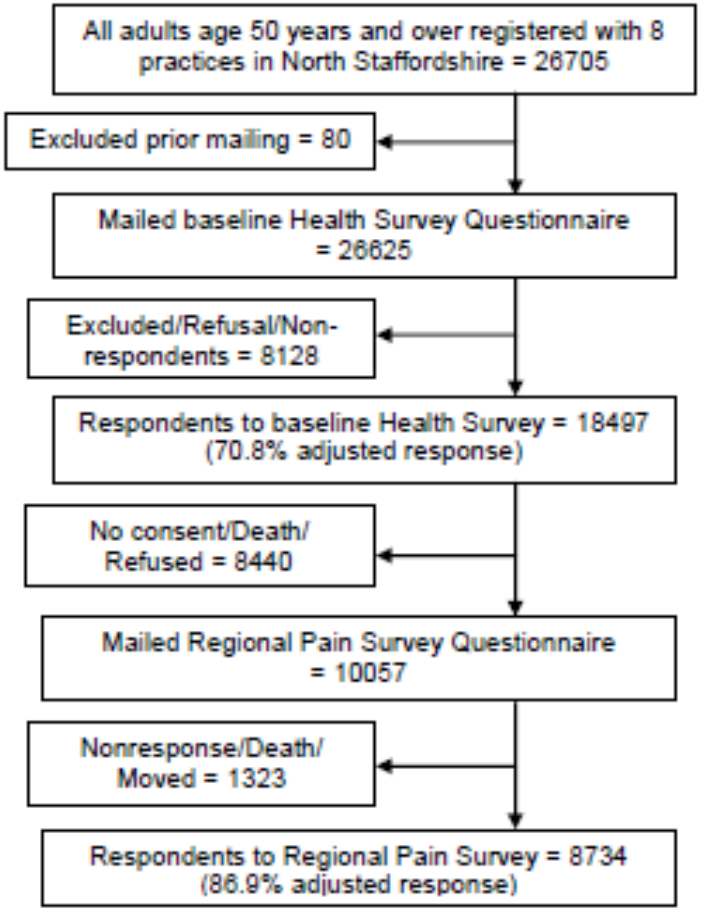
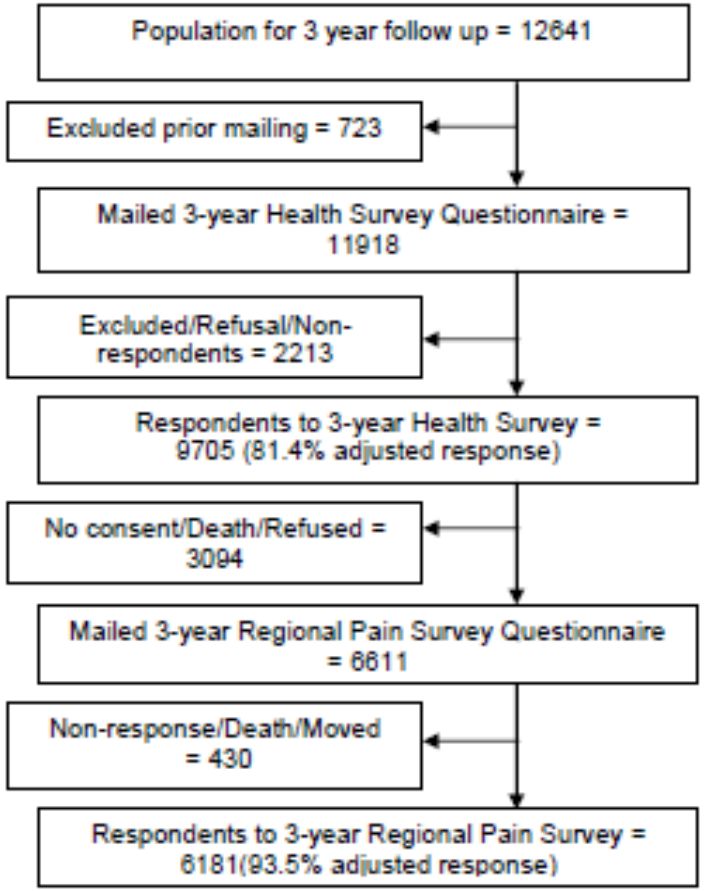


Figure 4.1b. Flowchart of recruitment into the NorStOP cohorts at 3 year follow up.



4.2 Comparison of responders and non-responders at baseline and 3 year follow up

Table 4.1a shows the age and gender distribution of the NorStOP cohorts amongst responders and non-responders at baseline [*Muller 2010*]. When examined by gender, women aged 60 to 79 years were more likely to respond than those in the lowest and highest age groups. The proportion of male responders was slightly lower in the lowest age group (33.5% versus 48.8% male non-responders). It was not possible to examine if responders and non-responders differ with respect to other characteristics as these data was not available for analysis.

The distribution of age and gender of the NorStOP cohorts between responders and non-responders at 3 years follow up is illustrated in Table 4.1b below. Similar gender distributions were observed for both responders and non-responders at 3 years follow up. However, for both men and women those aged 60 to 79 years were more likely to respond than those in the lowest and highest age groups. The proportion of responders and non-responders at 3 years follow up were similar for other baseline characteristics including socio-demographic (BMI, anxiety, depression and social class), pain (hip, knee, hand and foot) and physical function variables.

Table 4.1a Age and gender distribution of responders and non-responders of the NorStOP 1, 2 and 3 participants at baseline

	Responders (%)		Non Responders[a] (%)	
Variable	**Female**	**Male**	**Female**	**Male**
50 to 59 years	31	33.5	36.2	48.8
60 to 69 years	29.1	32.8	22.3	27.5
70 to 79 years	25.6	24.2	19	15.4
80 years plus	14.3	9.4	22.4	8.4

Source: Muller S. 2010 PhD thesis

a - Non-responders included people who did not return a questionnaire, returned a blank questionnaire and those who contacted the Research Centre to say they no longer want to take part in the study, but were eligible to be included.

Table 4.1b Age and gender distribution of responders and non-responders of the NorStOP 1, 2 and 3 participants at 3 years follow up

	Responders n (%)		Non Responders n (%)	
Variable	**Female**	**Male**	**Female**	**Male**
50 to 59 years	1979 (37)	1578 (36)	506 (42)	513 (48)
60 to 69 years	1749 (33)	1600 (36)	300 (25)	300 (28)
70 to 79 years	1203 (23)	997 (23)	246 (20)	186 (17)
80 years plus	375 (7)	224 (5)	152 (13)	77 (7)
Total	5306	4399	1204	1076

4.3 Description of baseline characteristics of participants among binary categories of pain and functional limitation.

Description of baseline variables for participants according to the binary 3 year follow up outcome categories of severe pain and functional limitation are shown in appendix 2. Only the variables that were selected in the final Poisson models for either pain or functional limitation are presented for the purpose of simplicity. The full list of variables used to derive the models has been described in chapter 3 section 3.3.4. The variables selected in

the logistic regression model but not in the Poisson regression model and vice versa are described in section 4.6 below.

71% of the participants had severe pain at three year follow-up whilst less than one-third (29%) were classified as having mild or no pain. There was no difference in age between participants with mild or no and severe pain (median age 63 versus 64 years) and a slightly higher proportion of the women suffered severe pain at three years.

The proportion of participants with poor functional outcome was 47% at three years follow-up. They were slightly older than those with good physical function (median age 66 versus 62 years), and a slightly larger proportion were female.

4.4 Baseline predictors of severe pain at three years in the final Poisson regression model.

Table 4.2 shows the baseline predictors significantly linked with severe pain at 3 years in the final multivariable Poisson regression model with their respective estimates of IRR (95% CI), unadjusted PAR (95% CI), adjusted PAR (95% CI) and unadjusted NNT (95% CI). The complete case analysis was based on 1643 participants. The baseline predictors independently associated with increased risk of severe pain at 3 years were knee pain in last year, high baseline score for WOMAC knee pain, poor physical function (SF-36) at baseline, hand pain in the last year, no full time education after school, obesity, poor AUSCAN function, poor AUSCAN pain, presence of hip pain in the last year, increased anxiety scores, no access to advice/help with income, raised blood pressure and front right

foot pain. In order to visualize the IRRs with (95%CI) more clearly, they are displayed in figure 4.2a.

The predictors associated with decreased risk of severe pain were no or few days use of natural remedies in last 4 weeks, waiting several days when ill before consulting the GP, not attending club/church/social events and no/few days use of painkillers in last 4 weeks. Reversing the reference categories of these predictors would change the direction of association to increased risk of severe pain – however, leaving the reference categories as they are is appropriate as the focus is placed on categories a priori assumed to be associated with poor outcome (i.e. supposedly vulnerable group). Note that in figure 4.2a and all such subsequent figures, a reference line of no effect (value = 1) has been placed on the x-axis.

Figure 4.3a below shows the estimates of adjusted PAR, unadjusted PAR and unadjusted NNT for only those predictors (in descending order from strongest effect size to the least) associated with increased risk of severe pain. Both adjusted and unadjusted PAR estimates were plotted to illustrate their similarity. The (PAR) proportion of severe pain in the population at three year follow-up that can be attributed to reporting knee pain (in the last year) at baseline was 16%. The NNT associated with the strongest predictor (knee pain at baseline) was 5, whilst poor physical function (SF-36) at baseline was associated with the smallest NNT of 4. The overall PAR attributed to all predictors associated with increased risk was 56% (95% CI: 49% to 62%).

The presentation of the epidemiological indicators focuses on predictors associated with increased risk of poor outcome, as these predictors will be most helpful when identifying

the most relevant predictors of severe pain and limitations in function. The PAR and NNT estimates for predictors associated with decreasing pain are less meaningful and less useful for clinical application even though they give an indication of the attributes that are associated with a reduced risk of long-term pain and functional limitation. A protective predictor would be associated with a number needed to harm (NNH) instead of NNT, and would lead to a negative upper 95% confidence limit. In table 4.2 for example, the NNT with 95% CI estimated for participants who had no access to advice or help with income is 62 (20 to -54) and can be re-written as 62 (NNB 20 to ∞ to NNH 54) which implies that the NNT for one person to benefit (NNB) ranges from 20 to infinity, and the number treated for one person to be harmed would be 54.

4.5 Baseline predictors of functional limitation at three years in the final Poisson regression model

The predictors associated with poor function at 3 years in the final Poisson regression model are shown in table 4.3 and also displayed in figure 4.2b with the number of subjects used to derive the model being 1602 (complete case analysis). The results show that the combination of baseline predictors associated with increased risk of poor function at 3 years included: poor function (SF-36), poor physical component score (SF-12), being retired from work, reduced time or change in activity in the last year, no full time education after school, having a knee problem in the last year, poor WOMAC hip function, increased depression score, not going out for a walk in a week, obesity, raised blood pressure, presence of previous years shoulder pain, elbow pain, hip pain and foot pain.

Predictors of a reduced risk of poor functional outcome included: caring for others for a few days per week, not shopping or shopping only on a few days per week, disagreeing that the GP can do a lot to help with joint pain, disagreeing that one's health is unpredictable and not having reduced time/changed activities in last 4 weeks.

The estimates of the adjusted PAR, unadjusted PAR and unadjusted NNT for the predictors associated with increased risk of poor function at 3 years are displayed in figure 4.3b. In general, similar estimates were obtained for both the adjusted and unadjusted PARs. The estimates of adjusted PAR and unadjusted NNT for the strongest predictor (i.e. poor physical function at baseline) were 48% and 2 respectively. The total proportion of poor function at three years that can be attributed to all the baseline predictors associated with increased risk was 80% (95% CI: 76% to 84%).

Table 4.2: Final Poisson regression model for severe pain at three years

Variables	N	IRR(95% CI)	Adjusted PAR (95% CI)	Unadjusted PAR(95% CI)	Unadjusted NNT(95% CI)
Knee pain in last year					
No	893	1			
Yes	2638	1.31(1.19, 1.43)	16.3(13.6, 18.9)	18.6(5.0, 32.1)	5.1(4.3, 6.3)
WOMAC knee pain at Baseline					
Low(Good symptom)	1127	1			
High(Pood symptom)	2436	1.15(1.06, 1.24)	10.6(8.6, 12.5)	9.2(-1.4, 19.7)	6.18(5.1, 7.8)
Physical function (SF-36) score at baseline					
High (Good function)	1907	1			
Low (Poor function)	1656	1.14(1.07, 1.22)	7.0(5.9, 8.0)	10.8(6.0, 15.6)	3.6(3.3 to 4.0)
Hand pain last year					
No	1097	1			
Yes	2429	1.12(1.05, 1.20)	8.3(6.5, 10.0)	7.7(-4.0, 19.3)	7.4(6.0, 9.9)
Go onto full time education after school					
Yes	457	1			
No	3042	1.12(1.01, 1.23)	7.9(4.7, 11.0)	9.1(-38.3, 56.5)	8.6(6.1, 14.6)
BMI					
Normal weight	1111	1			
Obese	815	1.11(1.05, 1.17)	2.7(2.1, 3.2)	2.5(-9.8, 14.8)	5.7(4.7, 7.2)
AUSCAN Physical function					
Low(Better)	1273	1			
High(Poor)	2290	1.10(1.01, 1.21)	8.4(6.3, 10.5)	6.2(-1.5, 13.9)	4.8(4.1, 5.6)
AUSCAN pain					
Low(Better)	1229	1			
High(Poor)	2334	1.10(1.01, 1.20)	4.2(2.3, 6.1)	6.0(-2.7, 14.8)	5.5(4.7, 6.7)
Hip pain in last year					
No	1720	1			
Yes	1801	1.09(1.03, 1.16)	4.5(3.4, 5.5)	4.4(-1.0, 9.7)	6.9(5.7, 8.6)
HADS Anxiety					
Low (Little distress)	1601	1			
High (Most distress)	1962	1.09(1.02, 1.15)	3.5(2.4, 4.6)	4.5(-1.8, 10.8)	7.4(6.1, 9.5)
Access to advice or help with income					
Yes	2637	1			
No	853	1.08(1.01, 1.15)	1.1(1.0, 1.6)	1.9(-0.6, 4.3)	61.9(19.7, -54.2)

Table 4.2 continued

Variables	N	IRR(95% CI)	Adjusted PAR (95% CI)	Unadjusted PAR(95% CI)	Unadjusted NNT(95% CI)
Raised blood pressure					
No	2273	1			
Yes	1290	1.07(1.01, 1.13)	1.5(0.8, 2.2)	2.6(-0.9, 6.0)	12.5(9.1, 20.0)
Front right foot pain (man38)					
No	2537	1			
Yes	1026	1.06(1.01, 1.13)	1.8(1.3, 2.3)	1.8(-0.6, 4.2)	6.07(5.2, 7.4)
Natural remedies last 4 wks					
All of most days	1327	1			
Few or no days	1616	0.93(0.88, 0.99)	-	-	-
When one goes to the doctor when ill					
Immediate or wait for few days	1247	1			
Wait several days	1191	0.93(0.87, 0.99)	-	-	-
Go to a club, church or social event					
Most days in a week	220	1			
No day in a week	1215	0.92(0.87, 0.98)	-	-	-
Painkillers in last 4 wks					
All or most days	1592	1			
Few or no day	1137	0.81(0.75, 0.88)	-	-	-

Pearson goodness of fit chi-square test = 1152.055, p-value = 1.000
C-Statistic or Area under ROC = 0.783 (0.764 to 0.801)
Total number of subjects used to derive the model = 1643
N – Number of subjects

IRR(95% CI) – Incident rate ratio (95% Confidence Interval)

PAR (95% CI) – Population Attributable Risk (95% Confidence Interval)

NNT(95% CI) – Number Needed to Treat(95% Confidence Interval)

Man – Body manikin: a tool made up of 50 items that covers the whole body used to measure bodily pain.

Table 4.3: Final Poisson regression model for functional limitation at three years

Variables	N	IRR(95% CI)	Adjusted PAR (95% CI)	Unadjusted PAR(95% CI)	Unadjusted NNT(95% CI)
Physical function (SF-36) score at baseline					
High (Good)	1907	1			
Low (Poor)	1656	2.48(2.02, 3.05)	48.3(43.3, 52.8)	40.8(37.2, 44.3)	1.7(1.7, 1.8)
Physical component (SF-12) Score at baseline					
High (Good)	1952	1			
Low (Poor)	1611	1.44(1.24, 1.67)	20.8(17.3, 24.2)	16.7(13.2, 20.1)	2.1(1.9, 2.2)
Current employment Status					
Employed	898	1			
Retired	1890	1.39(1.18, 1.64)	19.4(15.3, 23.4)	17.6(10.6, 24.6)	3.3(3.0, 3.7)
Unemployed	670	1.27(1.07, 1.51)	7.4(5.7, 9.1)	4.9(-2.1, 11.9)	2.7(2.4, 3.0)
Compared to 12 months Ago, have you reduced time or change how You have done any Activity					
No, Not at all	803	1			
Yes, a lot	188	1.31(1.01, 1.70)	7.8(5.6, 9.9)	6.2(-0.2 to 12.6)	1.7(1.6, 1.8)
Yes, a little	1009	1.37(1.07, 1.74)	13.3(8.7, 17.6)	15.9(9.5 to 22.3)	3.5(3.1, 3.9)
Walks of two miles or more					
Most day in a week	294	1			
No day in a week	2030	1.28(1.10, 1.48)	14.5(7.1, 21.4)	13.8(-1.6, 29.3)	2.6(2.3, 3.1)
Go onto full time education after school					
Yes	457	1			
No	3042	1.27(1.07, 1.52)	14.2(8.1, 19.8)	3.5(-34.7, 41.6)	8.7(6.2, 14.7)
Knee problem last year					
No	927	1			
Yes	2598	1.22(1.60, 1.40)	14.4(10.2, 18.4)	14.0(5.1, 22.9)	4.5(3.7, 5.3)
WOMAC hip physical function at baseline					
Low	904	1			
High	2659	1.21(1.06, 1.37)	8.6(4.8, 12.3)	13.3(1.3, 25.3)	6.3(5.1, 8.1)

Table 4.3 continued

Variables	N	IRR(95% CI)	Adjusted PAR (95% CI)	Unadjusted PAR(95% CI)	Unadjusted NNT(95% CI)
HADS Depression					
Low	1457	1			
High	2106	1.17(1.03, 1.33)	5.8(2.9, 8.6)	9.0(4.0, 14.0)	3.6(3.2, 4.0)
Back right shoulder (man 7)					
No	2531	1			
Yes	1032	1.16(1.04, 1.30)	4.3(3.0, 5.7)	4.5(2.0, 7.0)	5.4(4.5, 6.7)
Go out for a walk					
Most day in a week	1184	1			
No day in a week	720	1.16(1.07, 1.27)	3.8(2.7, 4.9)	3.2(-4.0, 10.5)	2.6(2.3, 2.9)
Front left elbow (man 29)					
No	3226	1			
Yes	219	1.16(1.03, 1.31)	1.1(0.7, 1.5)	1.5(0.3, 2.7)	5.0(3.9, 6.9)
Trouble staying asleep					
Not at all	889	1			
On some nights	1659	1.16(1.05, 1.28)	5.5(2.9, 8.0)	7.1(-5.5, 19.6)	9.2(6.8, 14.4)
Cost of living					
Quite comfortable	521	1			
Strain	154	1.16(1.01, 1.32)	1.2(-1.4, 3.7)	0.7(-22.7, 24.0)	2.6(2.2, 3.4)
Hip pain last year					
No	1720	1			
Yes	1801	1.14(1.02, 1.27)	4.0(1.8, 6.1)	6.5(1.7, 11.4)	5.9(5.0, 7.3)
BMI					
Normal weight	1111	1			
Obese	815	1.11(1.02, 1.22)	2.9(1.8, 3.9)	2.6(-7.0, 12.3)	4.1(3.5, 5.0)
Raised blood pressure					
No	2273	1			
Yes	1290	1.11(1.02, 1.22)	3.6(2.4, 4.9)	4.0(0.8, 7.1)	6.2(5.1, 7.9)
Front right foot (man 38)					
No	2537	1			
Yes	1026	1.10(1.01, 1.21)	1.8(0.8, 2.8)	2.8(0.3, 5.4)	6.0(5.0, 7.7)
Front right hip (man 46)					
No	2812	1			
Yes	751	0.90(0.81, 0.99)	-	-	-
I have power to influence					
What happens in my life					
Agree and strongly agree	2154	1			
Neither agree or disagree	901	0.90(0.81, 0.99)	-	-	-

Table 4.3 continued

Variables	N	IRR(95% CI)	Adjusted PAR (95% CI)	Unadjusted PAR(95% CI)	Unadjusted NNT(95% CI)
Look after others					
Most day in a week	997	1			
No day in a week	1548	0.89(0.80, 0.99)	-	-	-
Doctors can do a lot to help people with joint Problems					
Agree and strongly agree	2391	1			
Disagree and strong disagree	331	0.86(0.74, 0.99)	-	-	-
Neither agree or disagree	809	0.86(0.77, 0.97)	-	-	-
Go shopping					
Most day in a week	644	1			
Few days in a week	2691	0.86(0.74, 0.99)	-	-	-
No day in a week	152	0.73(0.60, 0.90)	-	-	-
Trouble falling asleep					
Not at all	1188	1			
On some nights	1718	0.79(0.70, 0.89)	-	-	-
On most nights	590	0.86(0.75, 0.98)	-	-	-
Back neck (man 43)					
No	2380	1			
Yes	1183	0.82(0.73, 0.91)	-	-	-
My health is very Unpredictable					
Agree and strongly agree	1355	1			
Disagree and strong disagree	1092	0.82(0.71, 0.93)	-	-	-
In past 4 weeks, have you reduced time/change how you have done any activity because of health					
Most day in a week	652	1			
No day in a week	1276	0.79(0.65, 0.95)	-	-	-

Pearson goodness of fit chi-square test = 1291.019, p-value = 1.000
C-Statistic or Area under ROC = 0.884 (0.871 to 0.896)
Total number of subjects used to derive the model = 1602
N – Number of subjects

IRR(95% CI) – Incident Rate Ratio (95% Confidence Interval)

PAR (95% CI) – Population Attributable Risk (95% Confidence Interval)

NNT(95% CI) – Number Needed to Treat(95% Confidence Interval)

Man – Body manikin: a tool made up of 50 items that covers the whole body used to measure bodily pain.

Figure 4.2a. Predictors of severe pain at three years in the final Poisson regression model

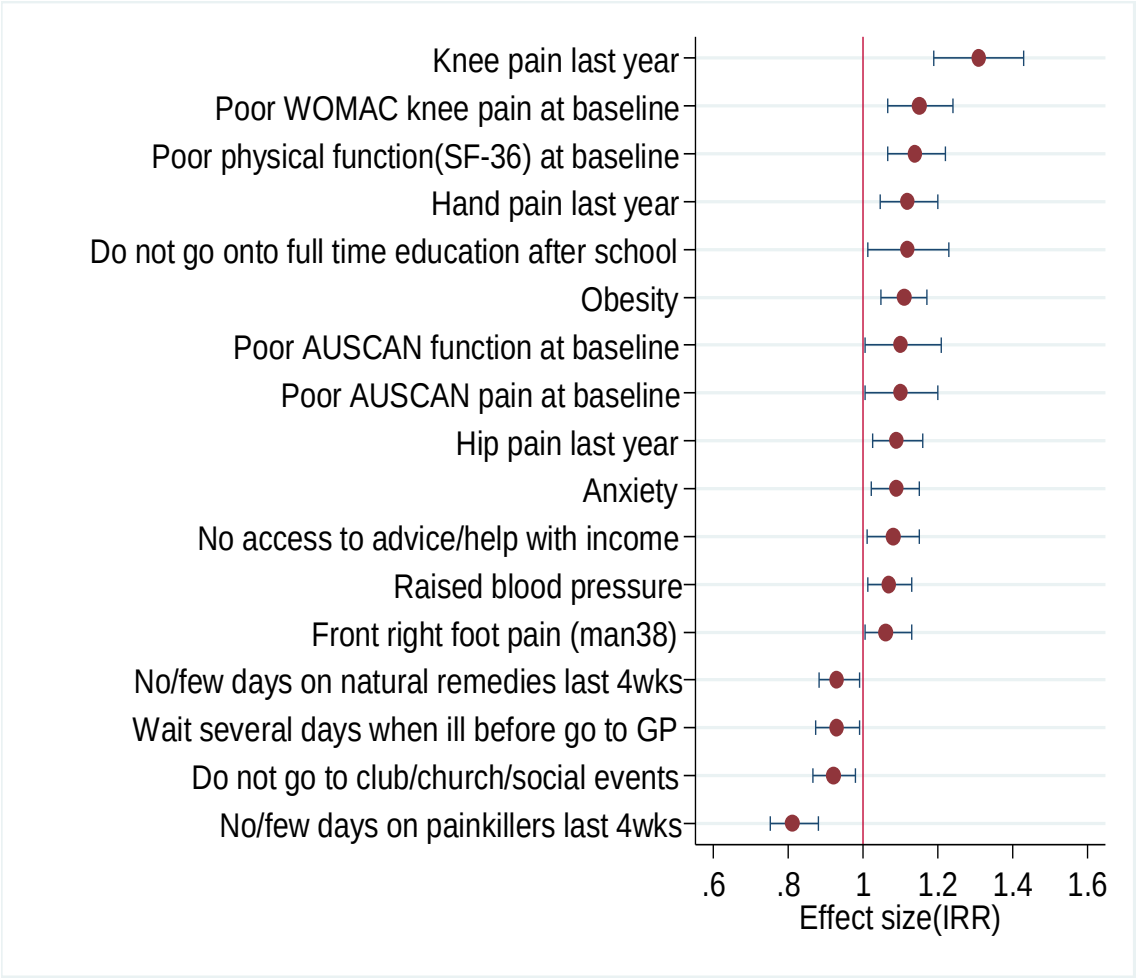


Figure 4.2b. Predictors of functional limitation at three years in the final Poisson regression model

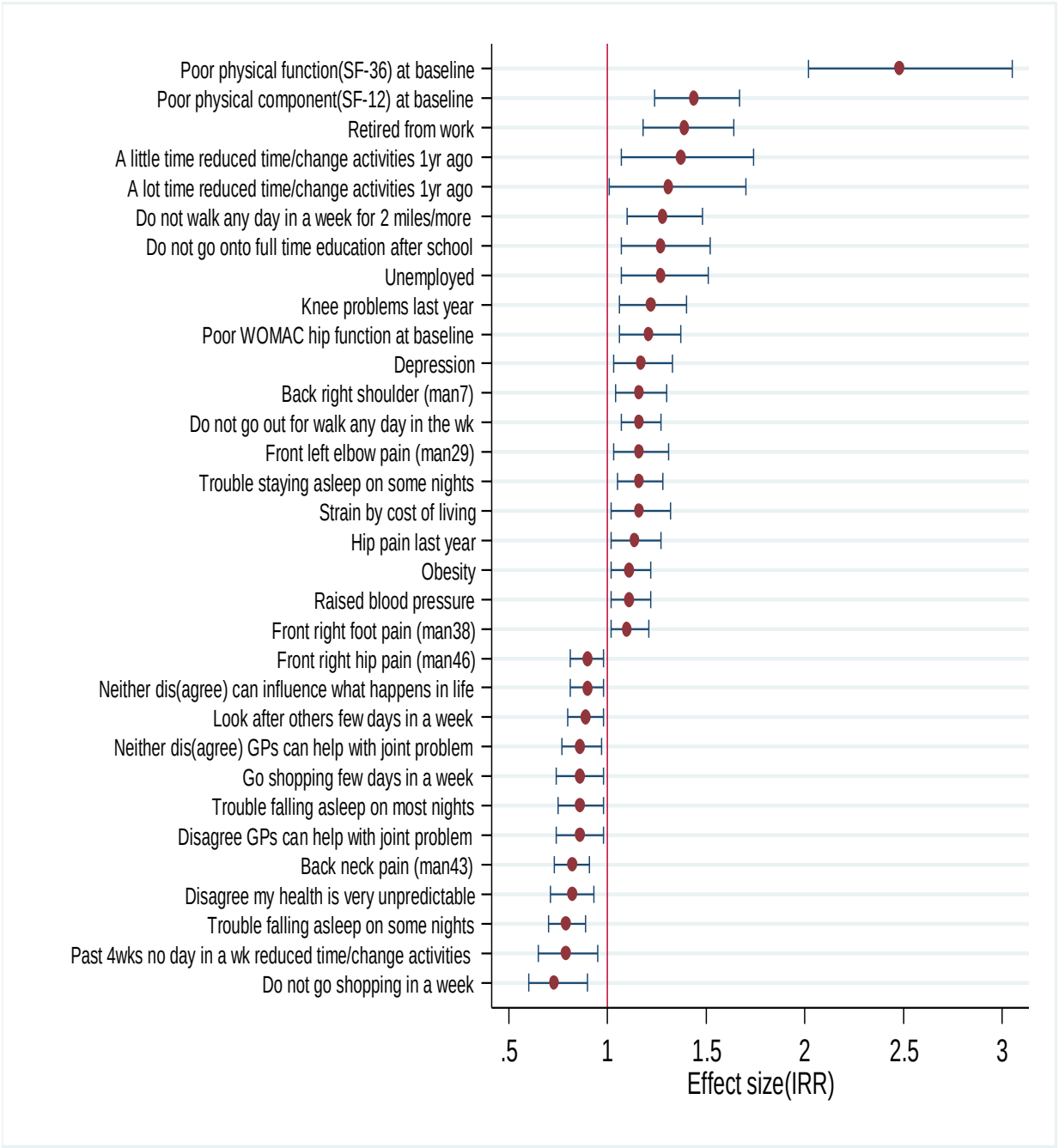


Figure 4.3a. Adjusted PAR, unadjusted PAR and unadjusted NNT for predictors associated with increase severe pain at three years in the final Poisson regression model

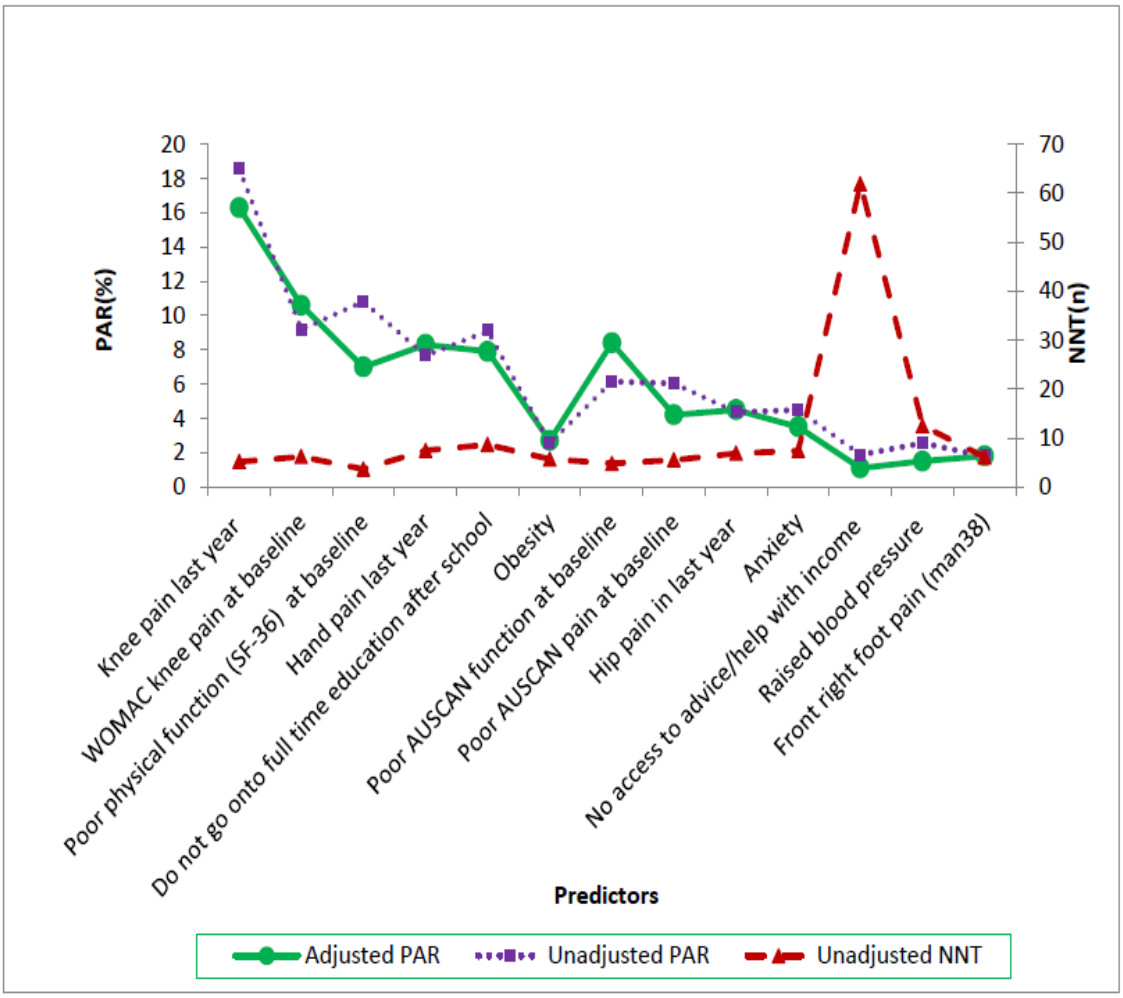
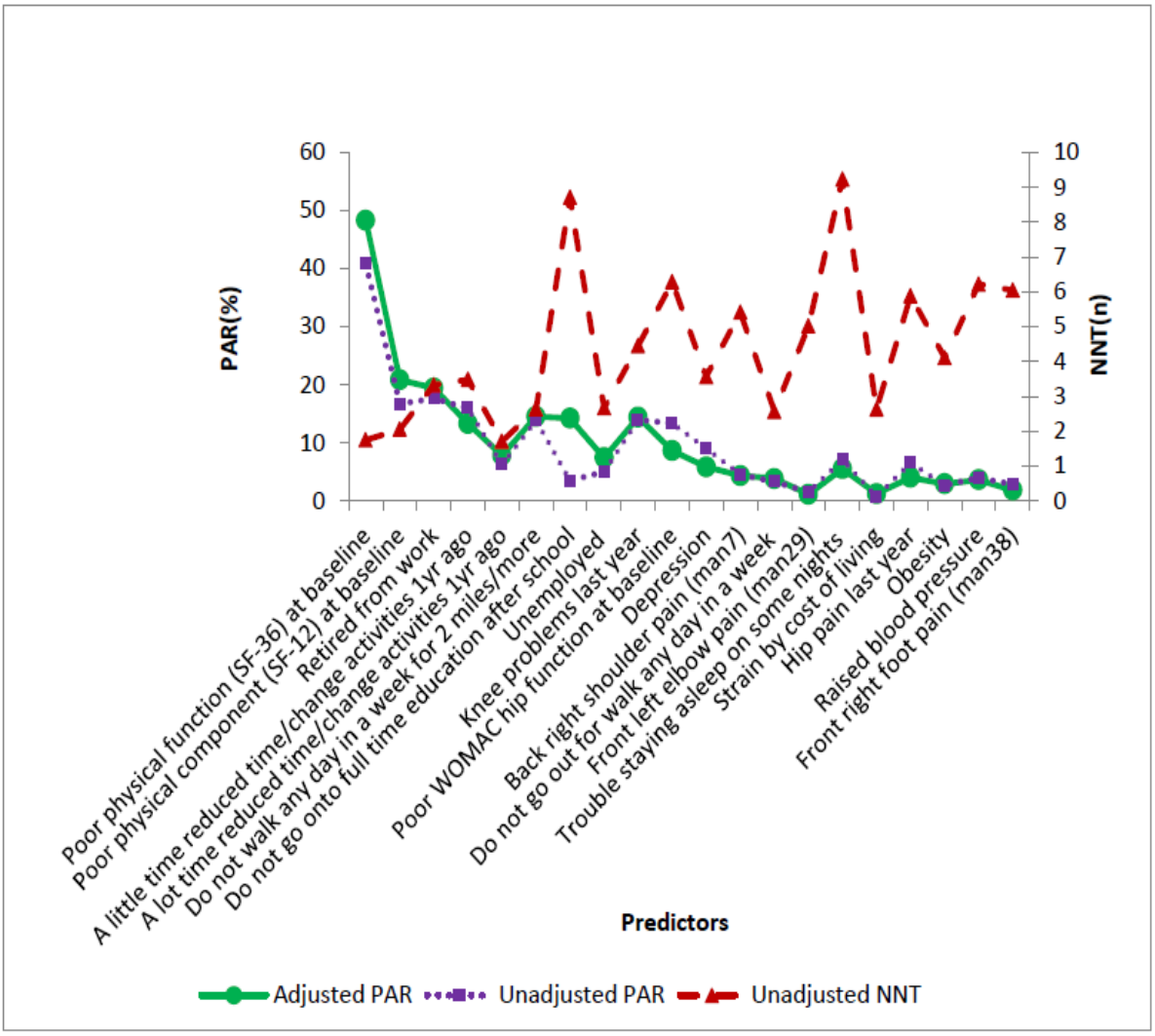


Figure 4.3b. Adjusted PAR, unadjusted PAR and unadjusted NNT of predictors associated with increase functional limitation at three years in the final Poisson regression model



4.6 Comparisons with logistic regression models

This section focuses on the comparison of the results of the logistic regression models with those of the Poisson regression models for both pain and functional limitation to examine if there are any differences between findings when using these different models.

Severe pain

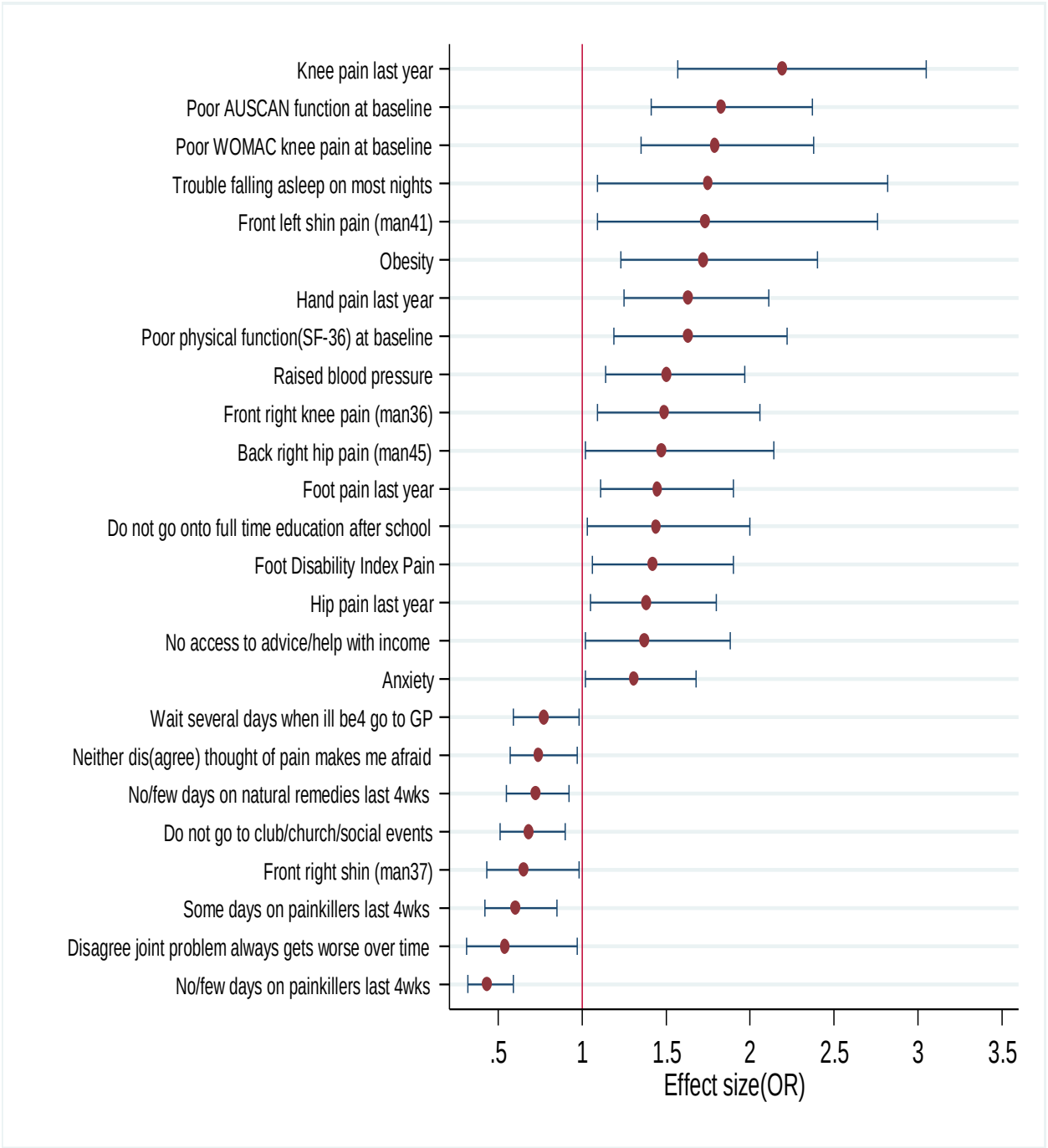
The set of predictors of severe pain which were selected in the final logistic regression model (table in appendix 3) were similar to that of the final Poisson regression model presented in section 4.4 except that the logistic model selected more predictors (25) than the Poisson model (17 predictors) and the OR estimates (as expected) were generally higher than the IRR estimates.

The additional predictors associated with increased odds of severe pain which were selected in the logistic regression model included pain in the front left shin, pain in the front right knee, pain in back right hip, foot pain in the last year, high foot pain disability index score and having trouble falling asleep on most nights (figure 4.4). These six additional predictors retained by the logistic regression model were not among the stronger predictors of outcome.

Two predictors associated with increased pain (pain in the front right foot and poor AUSCAN pain at baseline) were selected in the final Poisson model but not in the logistic regression model although they were not among the stronger predictors of outcome.

The top six predictors (in descending order) of severe pain in the final logistic regression model (figure 4.4 - i.e. having knee pain last year, poor AUSCAN function at baseline, WOMAC knee pain at baseline, having trouble falling asleep most nights, pain at front left shin and being obese) were similar to that in the final Poisson regression model for pain (figure 4.2a) with similar estimates of unadjusted PARs and NNTs (see figures in appendix 5a and figure 4.3a respectively). However, the logistic model showed slightly smaller adjusted PAR estimates compared to the Poisson model and this may have occurred because of the larger adjusted effects of the ORs estimates in the logistic regression model compared to the Poisson regression model. The number of participants used to develop both the logistic and Poisson regression models was the same (n=1643).

Figure 4.4. Predictors of severe pain at three years in the final logistic regression model



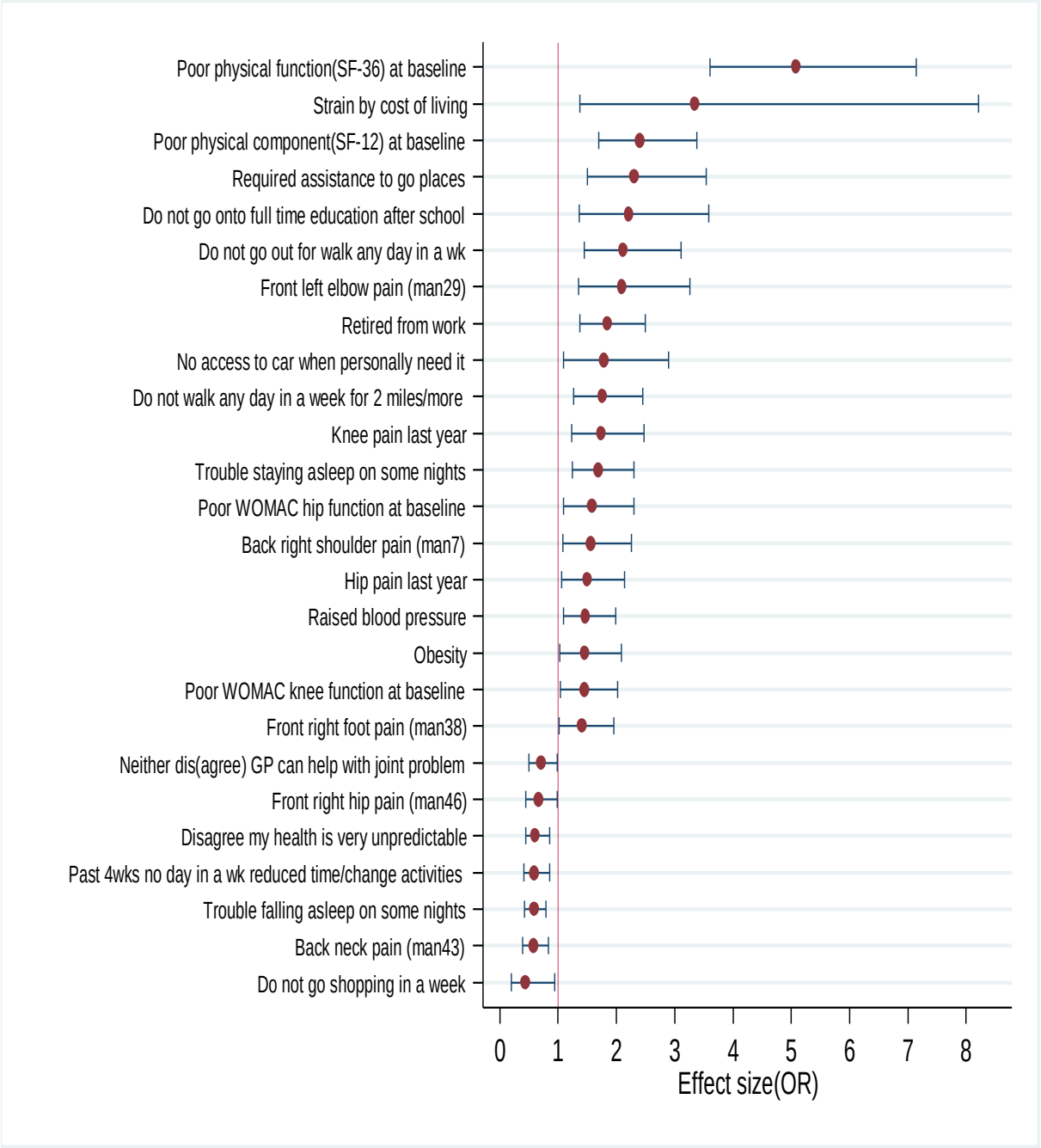
Pearson goodness of fit chi-square test = 2582.49, p-value = 0.686
C-Statistic (95% CI): = 0.793 (0.775 to 0.811)

Functional limitation

The set of predictors for poor physical function at 3 years identified in the final logistic regression model (table in appendix 4) is comparable to that of the final Poisson regression model presented in section 4.5. However, the logistic model selected fewer prognostic predictors (26) than the Poisson model (32). Both models selected only a few predictors which the other failed to select, and these were not among the predictors most strongly associated with poor outcome.

In both logistic and Poisson models, poor baseline physical function and poor physical component score (SF12) at baseline were among the 6 strongest predictors of functional limitation at three years (figure 4.5 and figure 4.2b respectively). The adjusted PAR estimates for poor physical function (29%) and poor physical component score (13%) at baseline in the logistic model (figure in appendix 5b) were slightly lower than those in the Poisson regression model (48% and 21% respectively) (figure 4.3b) and this may be attributed to larger adjustment of the effects for the ORs in the logistic regression model compared to the Poisson regression model. However, their NNT estimates were similar with the lowest NNT (1.7) being estimated for poor physical function in both models.

Figure 4.5. Predictors of poor physical function at three years in the final logistic regression model



Pearson goodness of fit chi-square = 2961.69, p-value = 0.058
C-Statistic (95% CI): = 0.885 (0.872 to 0.897)

4.7 Goodness of fit of the models

Both Poisson and logistic regression models for severe pain and functional limitation outcomes demonstrated good fit to the data, with p-values > 0.05 as shown in table 4.4 below (also reported as footnotes of the appropriate model tables and graphs presented above.) However, the estimates of both AIC and BIC for the logistic regression models were small for both pain and functional limitation outcomes compared to the Poisson regression models which indicates that the logistic regression models fit the data slightly better than the Poisson regression models. The pictorial presentation (calibration plots) of the goodness-of-fit of the models also provided evidence that the logistic models fitted the data slightly better than the Poisson models for both pain and functional limitation (see figures 4.6a to 4.6d) and this may be because the latter model appears not to work very well with counts of zeros and ones as defined by the binary outcome measures although robust variance estimator was used to adjust for this. Nevertheless, the Poisson models were preferred because of the reasons given in chapter three section 3.4.2.1 which are that it produces RR estimates and these were needed to estimate PARs and NNTs for the predictors.

Table 4.4 Estimates of Pearson goodness-of-fit statistics with their respective p-values, AICs and BICs for the Poisson and logistic regression models for the severe pain and functional limitation outcomes

Models	Outcome	Number of parameters	LL estimates	AIC	BIC	Pearson GOF statistic	Pearson GOF p-value
Poisson	Severe pain	17	-1506.1	3046.1	3066.8	1152.1	1.000
Logistic	Severe pain	25	-773.8	1597.6	1628.0	2582.5	0.686
Poisson	Functional limitation	32	-990.8	2045.6	2084.2	1291.0	1.000
Logistic	Functional limitation	26	-599.7	1251.3	1282.6	2961.7	0.058

LL – Loglikelihood
AIC – Akaike Information Criteria
BIC – Bayesian Information Criteria
GOF – Goodness of fit

Figure 4.6a. Calibration plots of the final Poisson regression model for severe pain outcome

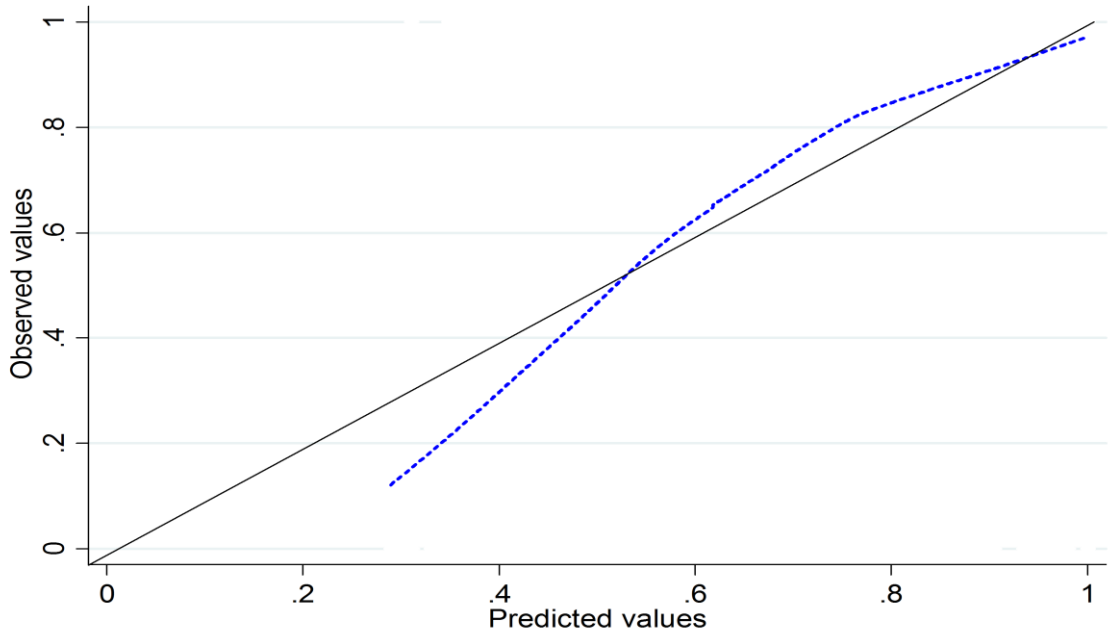


Figure 4.6b. Calibration plots of the final logistic regression model for severe pain outcome

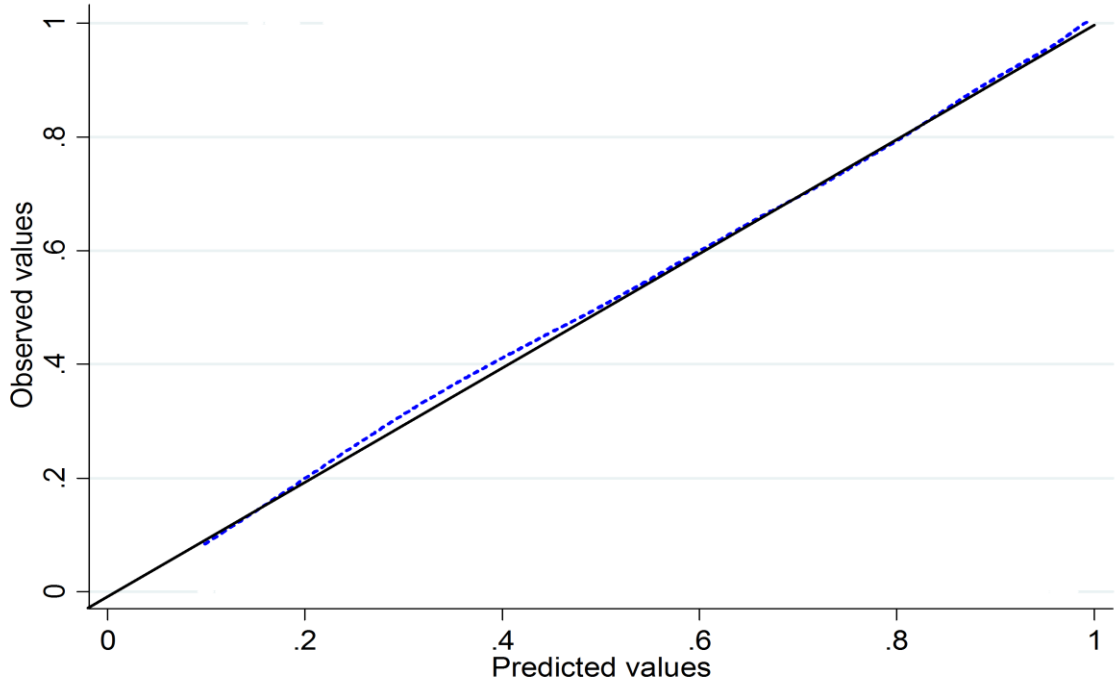


Figure 4.6c. Calibration plots of the final Poisson regression model for functional limitation outcome

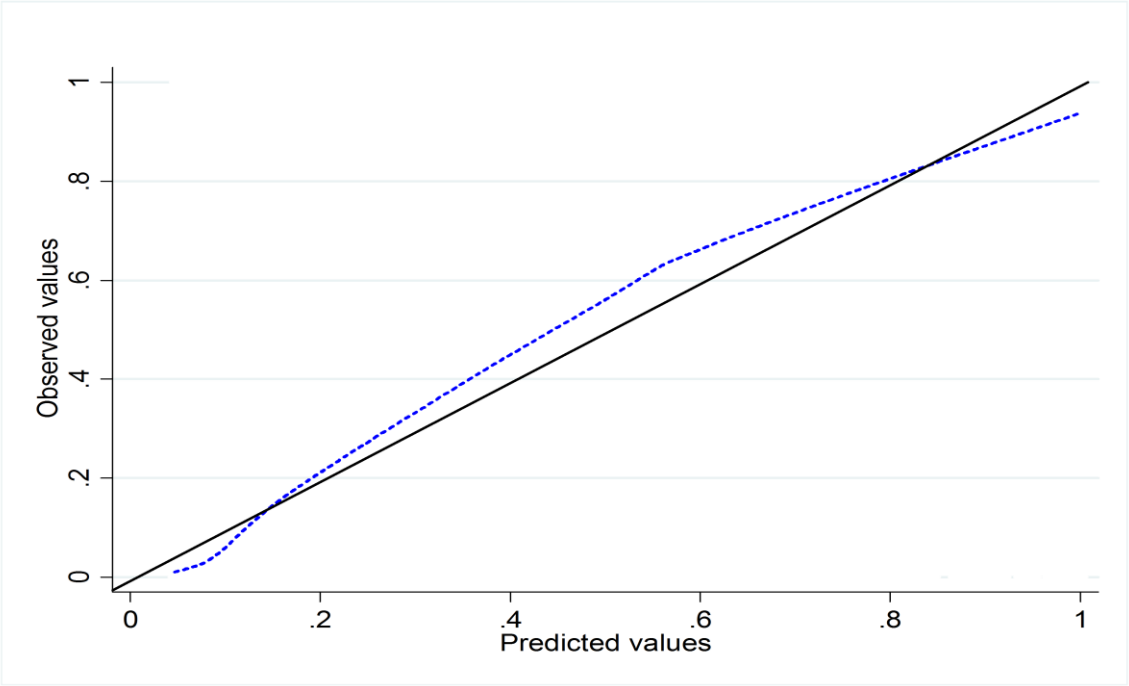
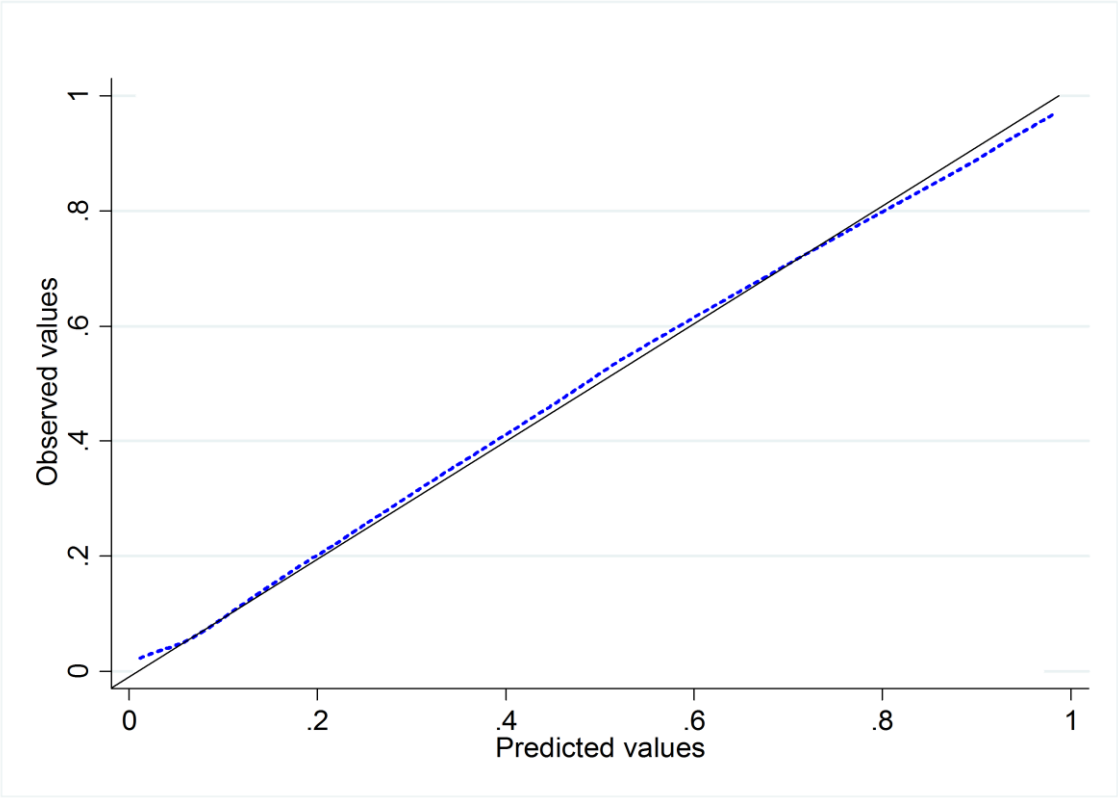


Figure 4.6d. Calibration plots of the final logistic regression model for functional limitation outcome



4.8 Internal validation of Poisson regression models

Severe pain

For the severe pain outcome analysis, the performance estimate of the final multivariable Poisson model (c-statistic = 0.783) was slightly higher than that based on the models derived from the 500 bootstrap samples (c-statistic = 0.765). The mean test sample model performance (c-statistic) was 0.811. With 500 bootstrap samples, the mean expected optimism value was -0.046 (i.e. 0.765 – 0.811) whilst the optimism corrected performance value was 0.829 (i.e. 0.783 - -0.046). The small estimate of optimism suggests that the final derived model has good internal validity (table 4.5).

Functional limitation

The findings of the Poisson model for functional limitation outcome showed that the performance estimate of the final multivariable model (c-statistic = 0.884) was higher than the average performance of the models derived from the bootstrap samples (c-statistic = 0.874). Mean test sample performance (c-statistic) was 0.898). Using 500 bootstrap samples the mean expected optimism value was -0.023 (i.e. 0.874 – 0.898) whilst the optimism corrected performance estimate was 0.907 (i.e. 0.884- -0.023). Since the optimism estimate was small, it suggests a small and insignificant under-estimation of the final derived model and hence good internal validity (table 4.5).

Table 4.5 Estimate of optimism in the Poisson regression models for severe pain and functional limitation at three years.

	Severe Pain	Poor Functional Limitation
	C-Statistic (95% CI)	**C-Statistic (95% CI)**
Apparent Performance[a]		
C-Statistic	0.783 (0.764 to 0.801)	0.884 (0.871 to 0.896)
Bootstrap Performance[b]		
C-Statistic	0.765 (0.764 to 0.766)	0.874 (0.873 to 0.875)
Test Performance[c]		
C-Statistic	0.811 (0.809 to 0.813)	0.898 (0.897 to 0.899)
Expected Optimism[d]		
C-Statistic	-0.046 (-0.048 to -0.045)	-0.023 (-0.024 to -0.022)
Optimism Corrected Performance[e]		
C-Statistic	0.829 (0.812 to 0.846)	0.907 (0.895 to 0.918)

CI – Confidence Interval
a – Model performance from the original (entire) sample (Final model)
b – Model performance from the bootstrap samples (Bootstrap model)
c – Model performance from the test sample (Bootstrap estimates applied to original sample)
d – Difference between bootstrap performance and test performance
e – Difference between apparent performance and optimism

Comparison with logistic model

The performance and optimism estimates for severe pain outcome based on the logistic models are shown in table 4.6 below. The performance estimate of the multivariable logistic model for severe pain (apparent c-statistic) was 0.793. Using 500 bootstrap samples, the mean expected optimism estimate (i.e. mean difference between bootstrap and test performance: 0.774 – 0.825) was -0.051 and the optimism corrected performance estimate was 0.844 (i.e. 0.793- -0.051). Since the value of optimism is small, it can be concluded that the model has good internal validity.

The estimates of the model performance and optimism for the functional limitation models are also shown in table 4.6 below. It can be seen from the table that the performance for the final multivariable logistic model (apparent c-statistic) was 0.885. Based on 500 bootstrap

samples, the expected optimism estimate was -0.032 (bootstrap performance: 0.872 – test performance: 0.904) whilst the optimism corrected performance estimate is 0.917 (i.e. 0.885 - -0.032). It can therefore be concluded that the final derived model has good internal validity.

Table 4.6 Estimate of optimism in the logistic regression model for severe pain and functional limitation at three years

	Severe Pain	Poor Functional Limitation
	C-Statistic (95% CI)	**C-Statistic (95% CI)**
Apparent Performance[a]		
C-Statistic	0.793 (0.775 to 0.811)	0.885 (0.872 to 0.897)
Bootstrap Performance[b]		
C-Statistic	0.774 (0.773 to 0.775)	0.872 (0.871 to 0.873)
Test Performance[c]		
C-Statistic	0.825 (0.823 to 0.826)	0.904 (0.903 to 0.905)
Expected Optimism[d]		
C-Statistic	-0.051 (-0.052 to -0.050)	-0.032 (-0.033 to -0.031)
Optimism Corrected Performance[e]		
C-Statistic	0.844 (0.827 to 0.861)	0.917 (0.905 to 0.928)

CI – Confidence Interval
a – Model performance from the original (entire) sample (Final model)
b – Model performance from the bootstrap samples (Bootstrap model)
c – Model performance from the test sample (Bootstrap estimates applied to entire sample)
d – Difference between bootstrap performance and test performance
e – Difference between apparent performance and optimism

In summary, the performance estimates of the Poisson regression models were similar to those of the logistic regression models for both the pain and functional limitation outcomes.

4.9 Selection of the most relevant predictors for severe pain in people with OA

In order to identify the most relevant predictors of outcome at three years (based on the Poisson regression model), the selection rule used by Smit et al [2006] was adopted. As explained in chapter 3 this rule helps to select predictors for which the highest possible health benefit (IRR and PAR) and the lowest possible effort and cost (NNT) can be achieved if interventions were to be fully effective.

The ranks of the effect sizes (IRRs), unadjusted PARs (in descending order from the strongest to the least) and of unadjusted NNTs (in ascending order from low to high) of the predictors in the Poisson regression model are shown in table 4.7 below. The unadjusted PARs estimates were similar to the adjusted PARs estimates and as such the former estimates were used in this study for consistency purposes since adjusted NNTs were not available. In this study IRRs and PARs were weighted more than the NNTs as the NNTs were generally similar for the predictors.

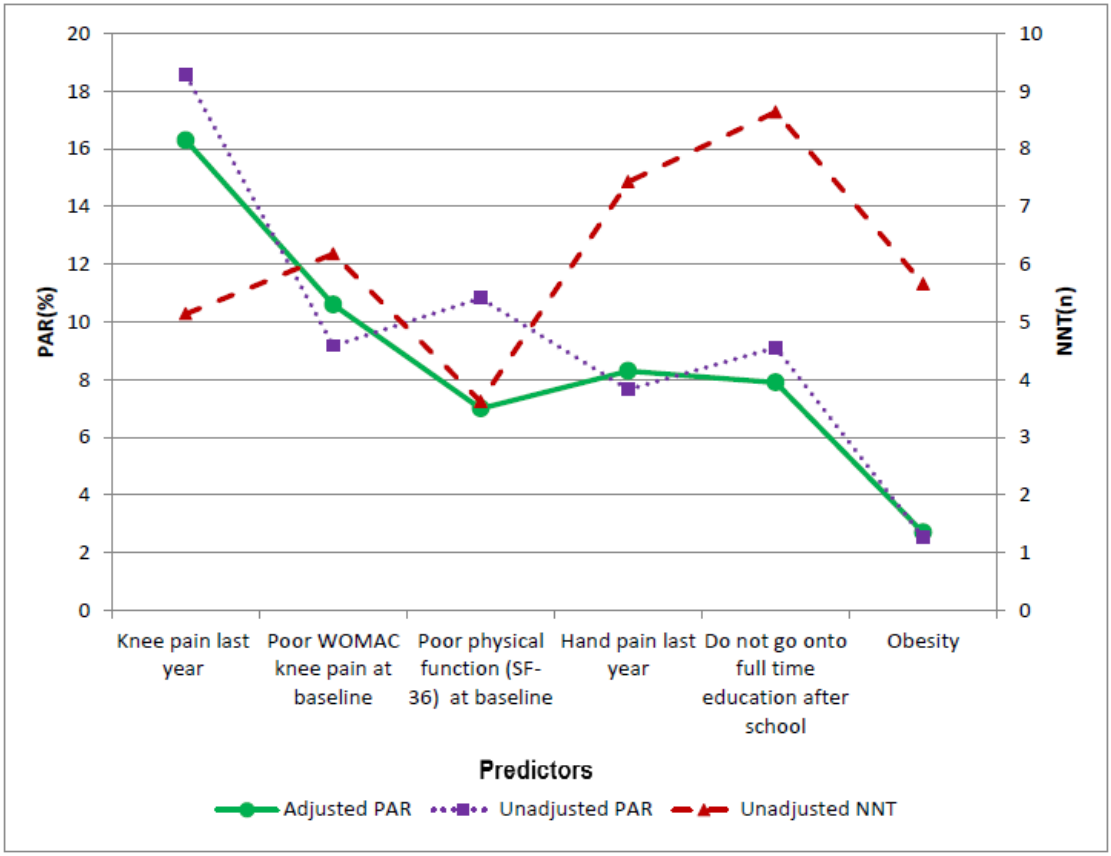
After applying the rule to the predictors in the final Poisson regression model (table 4.2) which was considered the preferred and primary model in this study, having knee pain in the last year, poor WOMAC knee pain, poor physical function (SF-36), hand pain in last year, not attending full time education after school and obesity were selected as the most important predictors for long-term severe pain (figure 4.7). The predictive performance based on this model of only six important predictors was (apparent c-statistic) 0.748 (95% CI; 0.730 to 0.766) compared to 0.783 (95% CI; 0.764 to 0.801) for the final, optimal Poisson regression model.

The unadjusted PAR estimates for previous year knee pain, WOMAC knee pain, poor physical function, hand pain last year, not attending full time education after school and obesity were 19%, 9%, 11%, 8%, 9% and 3% respectively whilst their unadjusted NNT estimates were 5, 6, 4, 7, 9 and 6 respectively (figure 4.7).

Table 4.7 Summary of ranks of predictors in final Poisson regression model for severe pain

Variables	Poisson Regression Model		
	Adjusted IRR	Unadjusted PAR	Unadjusted NNT
Knee pain last year	1	1	3
WOMAC knee pain at baseline	2	3	7
Poor physical function (SF-36) at baseline	3	2	1
Hand pain last year	4	5	9
Do not go onto full time education after school	5	4	11
Obesity	6	11	5
Poor AUSCAN function at baseline	7	6	2
Poor AUSCAN pain at baseline	8	7	4
Hip pain in last year	9	9	8
Anxiety	10	8	10
No access to advice/help with income	11	12	13
Raised blood pressure	12	10	12
Front right foot pain (man38)	13	13	6

Figure 4.7. Adjusted PAR, unadjusted PAR and unadjusted NNT for the top six predictors of severe pain at three years in the final Poisson regression model



4.10 Selection of the most relevant predictors for functional limitation in people with OA.

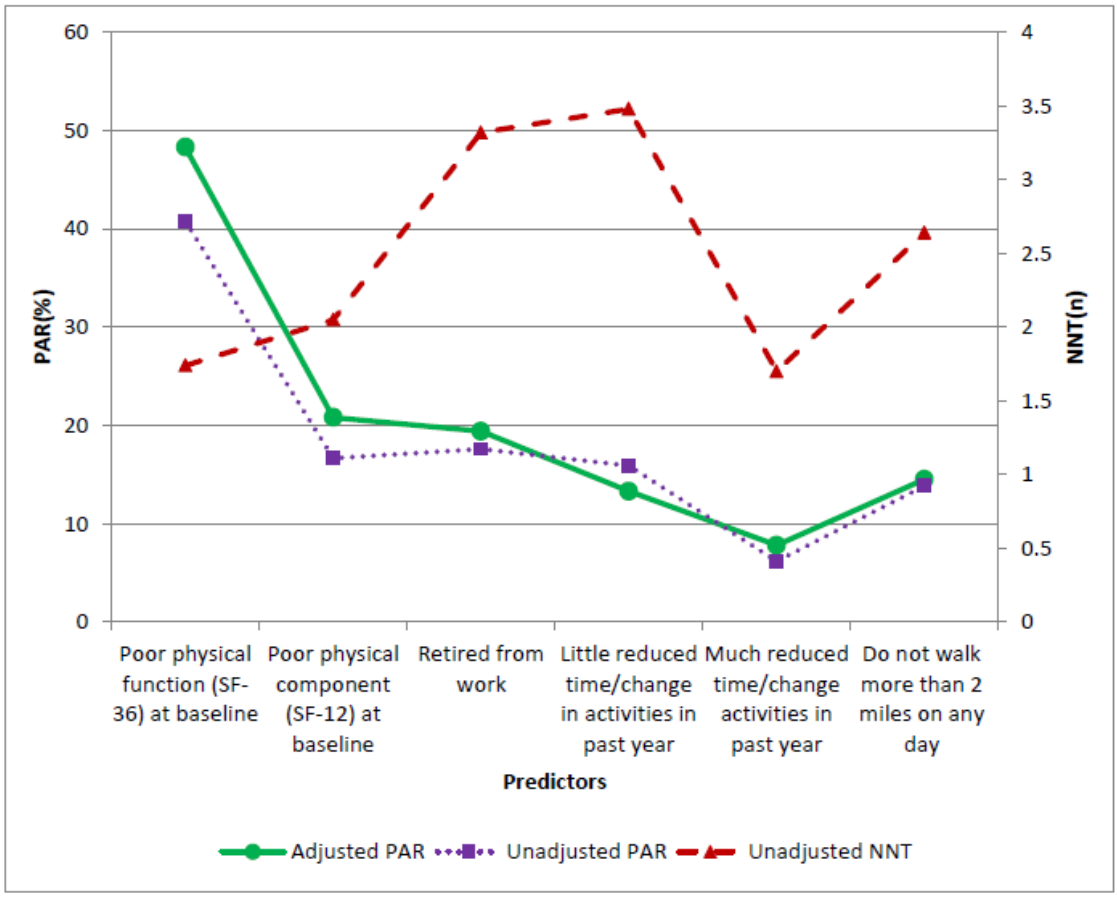
Table 4.8 below illustrates the ranks of the effect sizes, unadjusted PARs (in descending order high to low values) and of unadjusted NNTs (in ascending order of low to high values) of the predictors in the final Poisson regression model for poor function. The six most relevant predictors predicting poor physical function at 3 years were poor physical function (SF-36), poor physical component score (SF-12), being retired from work, reporting a little and a lot of reduction or change in activities in the past year and not walking any day in a week for 2 miles or more (figure 4.8). Their unadjusted PAR

estimates were 41%, 17%, 18%, 16%, 6% and 14%; whereas their unadjusted NNT estimates are 2, 2, 3, 3, 2 and 3 respectively. The predictive performance (apparent c-statistic) of the model based on these six predictors was 0.707 (95% CI; 0.688 to 0.725) compared to 0.884 (95% CI; 0.871 to 0.896) for the final Poisson regression model.

Table 4.8. Summary of ranks of predictors in final Poisson regression models for functional limitation.

Variables	Poisson Regression Model		
	Adjusted IRR	**Unadjusted** PAR	**Unadjusted** NNT
Poor physical function (SF-36) at baseline	1	1	2
Poor physical component (SF-12) at baseline	2	3	3
Retired from work	3	2	8
A little time reduced time/change activities 1yr ago	4	4	9
A lot time reduced time/change activities 1yr ago	5	11	1
Do not walk any day in a week for 2 miles/more	6	6	6
Do not go onto full time education after school	7	15	19
Unemployed	8	12	7
Knee problems last year	9	5	12
Poor WOMAC hip function at baseline	10	7	18
Depression	11	8	10
Back right shoulder pain (man7)	12	13	14
Do not go out for walk any day in a week	13	16	4
Front left elbow pain (man29)	14	19	13
Trouble staying asleep on some nights	15	9	20
Strain by cost of living	16	20	5
Hip pain last year	17	10	15
Obesity	18	18	11
Raised blood pressure	19	14	17
Front right foot pain (man38)	20	17	16

Figure 4.8. Adjusted PAR, unadjusted PAR and unadjusted NNT for the top six predictors of poor functional limitation at three years in the final Poisson regression model



4.11 Sensitivity analyses - multiple imputation results

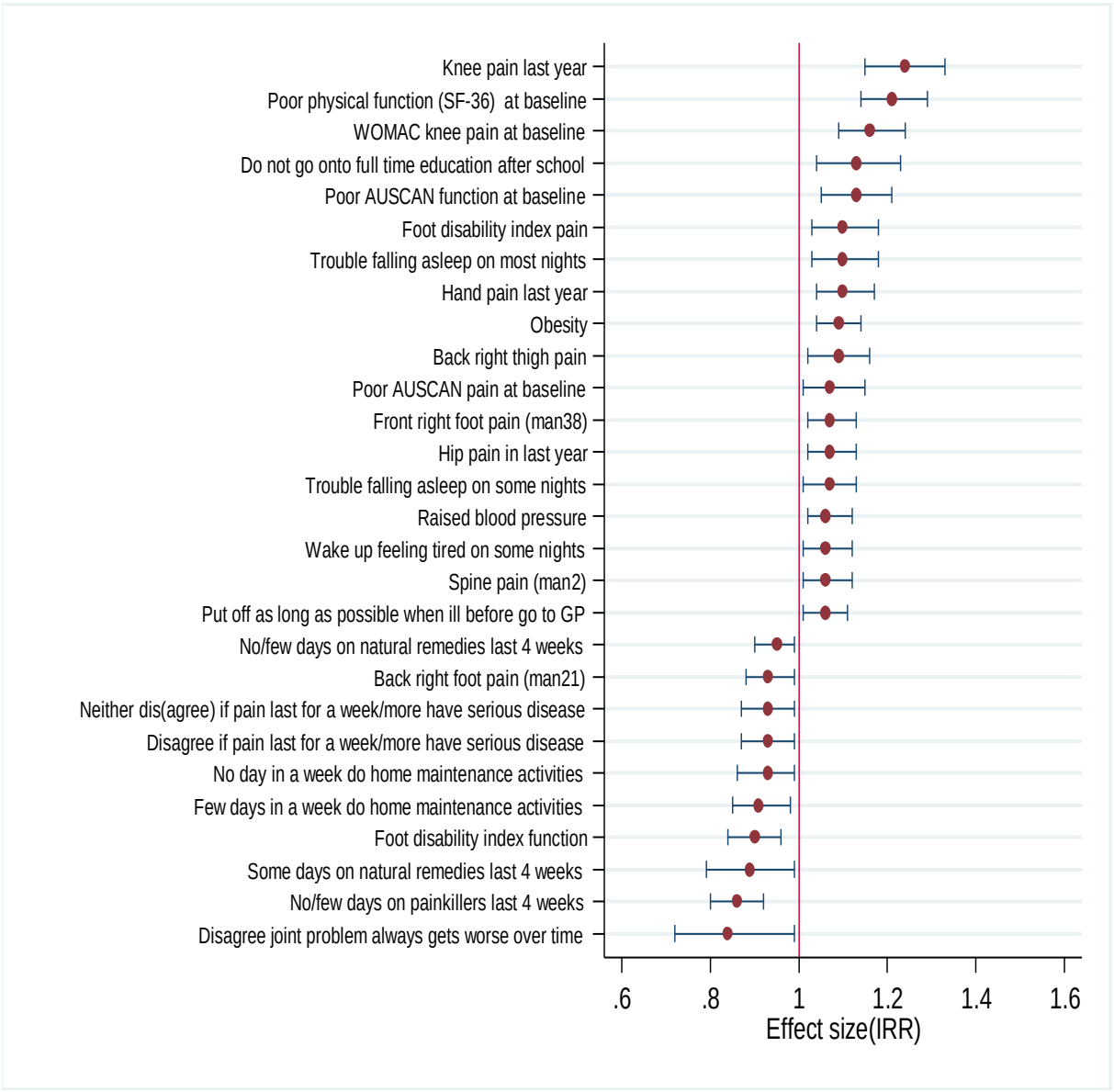
Missing values were imputed for 9 baseline variables with 3% or more missing values (i.e. go out to work – 3.9%, live alone – 4.2%, current employment status – 3%, trouble staying asleep – 3.4%, taking painkillers in last 4 weeks – 3%, applying creams/gels in last 4 weeks – 13.6%, taking natural remedies in last 4 weeks – 12.9%, participation restriction – 4.7% and social isolation – 18.6%). This resulted in an increase in the sample size used to derive the models from 1643 to 2510 (35% increase) and from 1602 to 2441 (34% rise) for the severe pain and functional limitation outcomes respectively. The models developed from the MI data were similar to those based on complete case data in terms of the

predictors retained, their effect estimates with 95% CIs and the top six predictors. Figures 4.9a to 4.9d show the predictors (listed in ascending order) with their effect estimates and 95% CIs selected in the respective Poisson and logistic regression models for severe pain and functional limitation outcomes with similar top six predictors. The number of predictors selected in the MI models was slightly higher than that of the unimputed data models except for the Poisson regression model for functional limitation outcome where equal numbers (33) of predictors (but slightly different predictors) were selected in both models. The effect estimates of the top six strongest predictors of the Poisson regression models for both the MI and unimputed datasets were similar for both outcomes. PARs and NNTs were not calculated for the MI models as there are no straight forward commands in STATA.

The full list of predictors selected in the respective Poisson and logistic regression models using the MI datasets but not in the unimputed data models and vice versa can be found in Appendix 6a to 6d. Because the MI datasets were larger than the unimputed datasets, this resulted in slightly more variables to be associated with outcome and hence caused the MI models to select slightly more predictors compared to the unimputed models. However, this did not lead to a major change in the six most important predictors for both outcomes.

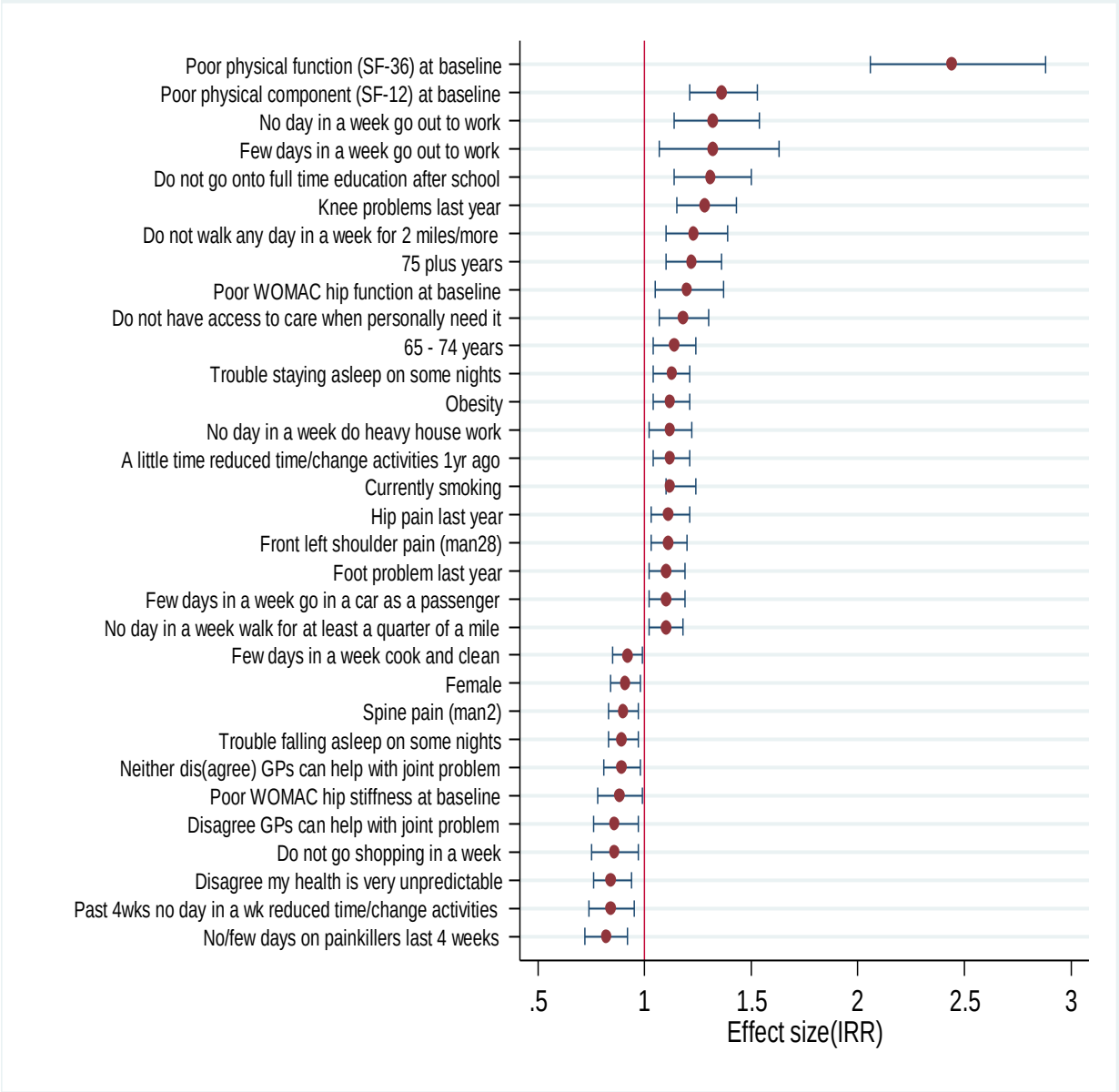
The c-statistic estimates were also similar for the respective Poisson and logistic regression models in both the MI and unimputed data models. The results of this analysis indicate that the missing data appeared to have little effect on the composition and performance of the prediction models for long-term outcomes of pain and functional limitations. (See table 4.9 below).

Figure 4.9a Predictors of severe pain at three years in the MI Poisson regression model



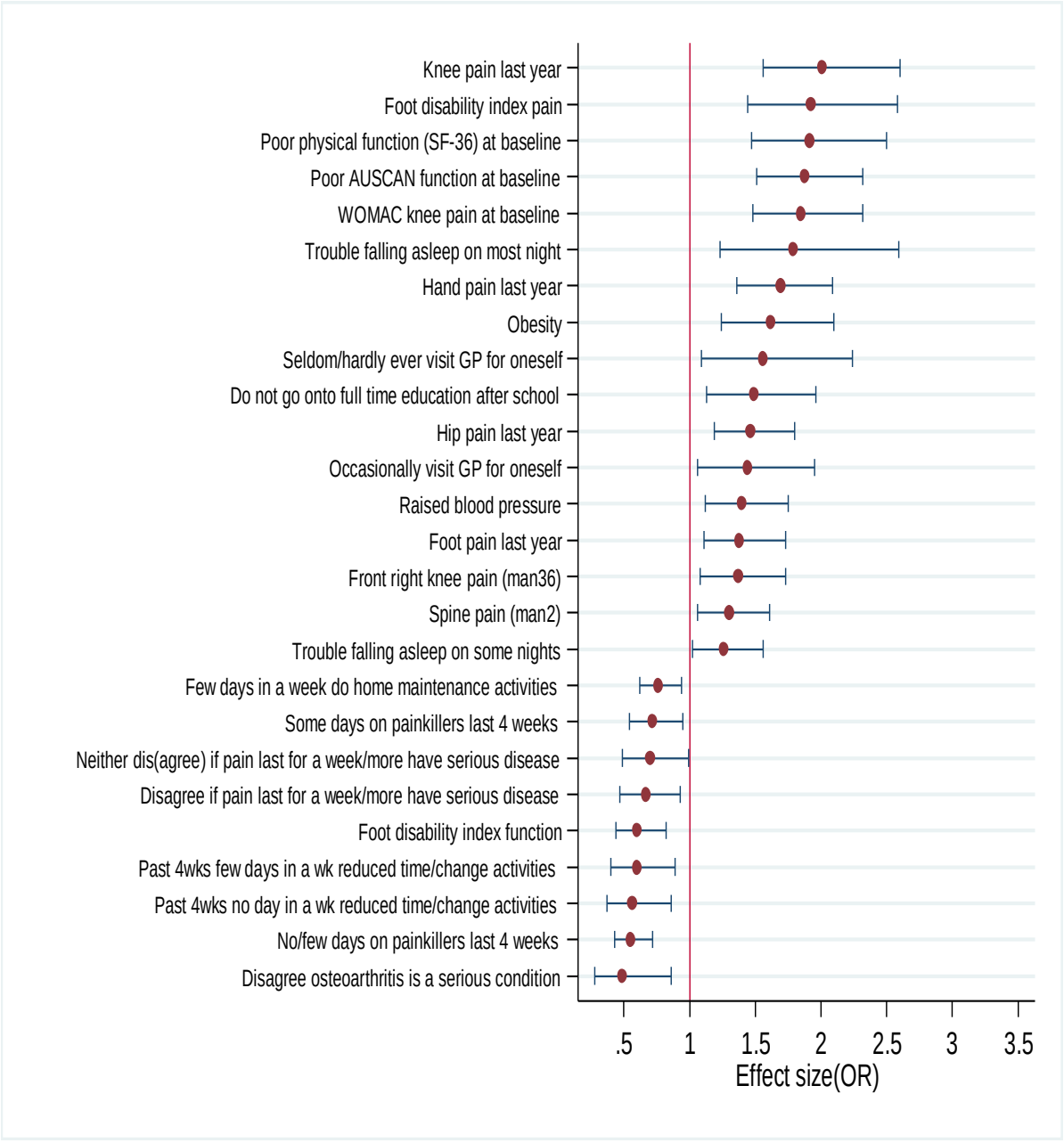
Number of subjects used to derive the model = 2510
C-Statistic or Area under ROC = 0.794 (0.778 to 0.811)

Figure 4.9b Predictors of poor function at three years in the MI Poisson regression model



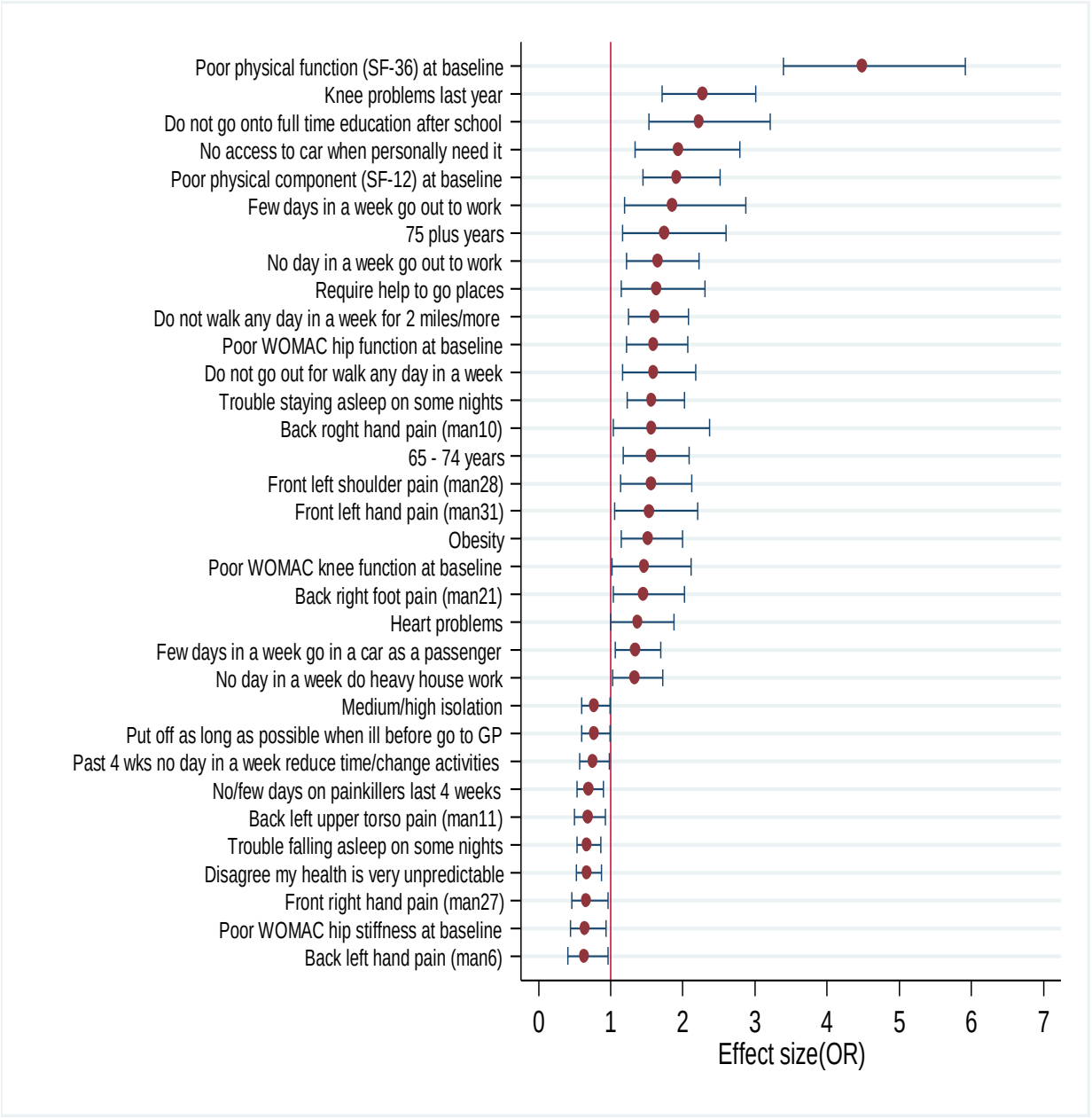
Number of subjects used to derive the model = 2441
C-Statistic or Area under ROC = 0.887 (0.875 to 0.899)

Figure 4.9c Predictors of severe pain at three years in the MI logistic regression model



Number of subjects used to derive the model = 2510
C-Statistic or Area under ROC = 0.798 (0.781 to 0.814)

Figure 4.9d Predictors of poor function at three years in the MI logistic regression model



Number of subjects used to derive the model = 2441
C-Statistic or Area under ROC = 0.894 (0.883 to 0.906)

Table 4.9. Comparison of the performance estimates for imputed and complete case Poisson and logistic models for pain and functional limitation outcomes

Models and outcomes	Subjects used: MI Versus Non-MI	No. of predictors Identified: MI Versus Non-MI	Imputed model C-Statistic (95% CI)	Complete case model C-Statistic (95% CI)
Poisson: severe pain	2510 Vs 1643	28 Vs 17	0.794 (0.778 to 0.811)	0.783 (0.764 to 0.801)
Logistic: severe pain	2510 Vs 1643	26 Vs 25	0.798 (0.781 to 0.814)	0.793 (0.775 to 0.811)
Poisson: functional limitation	2441 Vs 1602	32 Vs 32	0.887 (0.875 to 0.899)	0.884 (0.871 to 0.896)
Logistic: functional limitation	2441 Vs 1602	33 Vs 26	0.894 (0.883 to 0.906)	0.885 (0.872 to 0.897)

MI – Multiple imputation
Vs – Versus

4.12 Summary of chapter

In this chapter, I have presented the results of the prediction models for pain and functional limitation at 3 years, compared the findings of Poisson and logistic regression models, and selected the most relevant predictors. I have also illustrated the findings from analyses based on multiple imputed data sets and compared it with the findings from the complete case analysis. The meaning of these findings and the strengths and weaknesses of the methodology are discussed in the next chapter.

Chapter Five

Derivation of prediction models for OA – Discussion and conclusions

5.1 Summary of findings

The analysis of the combined NorStOP datasets in people aged 50 years and over with chronic joint pain in at least one joint site (hand, hip, knee or foot) showed that 71% of the participants suffered severe pain and 47% reported poor physical function at 3 years follow-up. Many baseline variables reflecting previous pain, physical, psychological and social variables were associated with severe pain and poor physical function at 3 years. The datasets also provided the opportunity to estimate maximum achievable (PAR) health gains and NNT which helped to identify a small set of relevant predictors that may help to identify patients with high risk of poor outcome.

The top six most relevant predictors of severe pain at three years were having knee pain in the previous year, poor WOMAC knee pain at baseline, poor physical function (SF-36) at baseline, hand pain in the previous year, not attending full time education after school and obesity whilst those for poor physical function were poor physical function (SF-36) at baseline, poor physical component score (SF-12) at baseline, being retired from work, reporting a little or a lot of reduction or change in activities in the past year and not walking 2 miles or more on any day in a week. Similar predictors were identified in both the Poisson and logistic regression models.

The tests of goodness of fit of the models for both severe pain and functional limitation outcomes confirmed that the models fitted the data well. Also, the performance estimates (c-statistic) of the models for both outcomes showed good internal validity, although model performance of the reduced models were somewhat lower than that of the full

model, and were higher than 0.70 (arbitrary cut-point for good performance) particularly for the models predicting poor functional limitation at three years.

5.2 Comparison with relevant findings from other studies

To my knowledge, no published study has developed optimal models of predictors of poor outcomes for OA regardless of the joints involved. However, OA is a condition that usually affects multiple joints, which will all impact on pain and function. Therefore, investigating outcomes of OA in the person may be more relevant than investigating the outcomes in a specific joint - hence the prediction modelling study presented in this thesis (in chapters 3 to 5) can be regarded as novel in its field. Available prediction studies [*Zhang et al 2011; Sa et al 2011; Thomas et al 2008; Jinks et al 2008; Mallen et al 2007; Topp et al 2000*] for OA examined specific joints rather than several joints and none estimated PAR and NNT for their predictors in their studies.

Since no study has examined predictors of poor outcomes of OA in any joint, the results of this study can only be compared with studies that investigated similar specific joints (i.e. hand, hip, knee or foot).

This study confirms the findings of previous population-based studies [*Yusuf et al 2011; Felson et al 2004*] by demonstrating that baseline status of pain and functional limitation are by far the strongest predictors of these outcomes at long-term follow-up. The adjusted PAR indicated that 16% of the risk of severe knee pain in the population could be attributed to knee pain at baseline, while this was 7% for baseline functional limitation. Dawson et al [2005] found high baseline pain score and number of painful hip or knee joints at baseline to be strongly related to severe hip or knee pain at 12 months in adults aged ≥ 65 years in a general population sample. Also, this study and previous studies

[*McAlindon et al 1993; Jordan et al 1997; Thomas et al 2008*] found that pain severity was associated with progression of functional disability of the knee. Although other studies used different measures of functional limitation, for example the function subscale of WOMAC instead of SF-36 scores [*Mallen et al 2007; Thomas et al 2008*], the results are similar. These findings confirm the fact that joint pain in older people represents a chronic pain problem, with the strongest predictors being baseline pain and functional limitation. The strong effect of pain on OA outcomes could also mean that, firstly, outcomes in OA are mainly driven by pain in the hip and/or knee joints which are mostly investigated in previous studies; secondly, most people have pain in multiple sites, which holds for the population investigated in this study but also for people included in other knee/hip OA studies; and thirdly, it may be that it does not matter where the pain is and that the consequences of pain are the same irrespective of the joint(s) affected.

Results of previous studies [*Zhang et al 2011; Sa et al 2011; Dawson et al 2005; Jinks et al 2008*] indicating a link between obesity and progression of knee pain were also confirmed by the findings of this study. This association could potentially be explained by reverse causality. Given that the participants already had pain 3 months or more prior to joining the study and the majority possibly over many years, it is possible that their pain caused them to reduce their physical activity levels which may subsequently have resulted in higher BMIs. The findings of this study however, suggest that obesity may not only be a predictor for the onset of joint pain and OA, but also indicate an increased risk in the progression of functional disability of the knee [*Felson et al 2004; Mallen et al 2007; Thomas et al 2008*].

Moreover, the results of this study adds to previous studies [*McAlindon et al 1993; van Baar et al 1998a; Sharma et al 2003*] by showing that socio-demographic factors such as education and retirement are associated with functional limitation at 3 years and that the former (education) is also associated with severe pain at 3 years.

The estimates of the predictive performances (c-statistics) obtained in this study demonstrated good, although not excellent, discriminative ability for all the models for both outcomes (i.e. severe pain and functional limitation at 3 years). The performance estimates of the models which selected the six most important predictors were lower compared to the full models, but still slightly higher compared to that of the study of Zhang et al [2011] (c-statistic = 0.69: 95%; 0.62 to 0.76) who developed risk prediction model which also comprised of six predictors (age, female, BMI, occupational risk, family history and knee injury) for onset of knee OA using a community-based cohort of 424 adults aged 40 to 79 years. The difference in the performance scores between this study and that of Zhang et al [2011] may perhaps be due to the fact that their sample was relatively small (n=424).

In contrast, the discriminative ability of the selected models in this study were slightly lower than that of Yusuf et al [2011] (c-statistics = 0.80; 95% CI: 0.76 to 0.94) which involved 117 adults aged 55 to 66 years with progressively severe knee or hip OA in the community and used clinical and radiographic predictors. Even though their sample is small and their model consisted of only three predictors, the slightly higher discriminative power they obtained may be a result of the composition of their population (which comprised of predominantly severe patients including those presenting for joint replacement) and the type of predictors used (including radiographic data; osteophytes and

joint space narrowing). Many factors may explain differences in results between studies, including for example, differences in the number and type of potential predictors used, the way predictors are measured, the method of variable selection procedure used and the various ways OA progression is defined by the researchers.

5.3 Strengths and limitations of this study

The NorStOP datasets provided a sample large enough to develop prediction models based on a large number of potential predictors of pain and functional limitation. The approach used to recruit the participants was standardized with high completion rates. Response rates to the baseline (71%) and 3 year follow up (81%) health survey questionnaires were adequate and only a small percentage of people actively opted out from the study for reasons such as refusal to consent for further data collection, transfer, death, etc. These response rates are comparable to rates in other population-based surveys [*Etter and Perneger 1997*]. Bowling [2002] recommended that a response rate greater than or equal to 75% is good, hence the 71% response rate achieved for the NorStOP cohorts can be said to be reasonable.

A few studies [*Montgomery et al 2010; Grobbee and Hoes 2007*] have argued that selection bias caused by loss to follow up should be investigated even when cohort studies have high follow up response rates. This is because loss to follow up is not always accidental and those with worse health status or more severe conditions may be the ones that may be lost to follow up [*Kristman et al 2004*]. However, one could also argue that participants whose disease status has improved over time may either not be interested in

research or feel that they would no longer be of importance to researchers and hence may decide not to continue with an on-going study.

The comparison of non-responders to responders at baseline showed that non-responders were more likely to be younger and male and may therefore also differ with respect to levels of pain, functional limitation and other socio-demographic characteristics [*Muller 2010*]. As this study examined predictors of pain and functional limitation among participants with pain for duration of three months or more at baseline, it is not likely that differences in levels of pain or functional limitations between baseline responders and non-responders has greatly affected the results. About one-third of the participants were lost to follow-up at 3 years and the main reasons were because participants did not consent for further contact or refused to continue with the study. The comparison between responders and non-responders to the 3 year RPS questionnaires shows that participants aged 60 to 79 years were more likely to respond whereas non-responders are likely to be younger or older. However, both responders and non-responders were similar with respect to socio-demographic, pain and functional limitation characteristics and as a result it unlikely that non response at 3 years will have had a major influence on the findings of this study.

Examination of missing data indicated that about half (47%) of the baseline variables had small proportions of missing data (average 3% per variable). Multiple imputation was performed on those variables with 3% or more missing data. It was computationally intensive to do this for all the variables with missing values and this could potentially lead to the estimation of biased effect sizes with imprecise confidence intervals. However, the findings from the analysis based on imputed data were similar to those based on complete

data in terms of the combination of predictors selected by the models, the point estimates and their level of precision, and the predictive performances for both Poisson and logistic regression method for both outcomes. Even though the models based on imputed data selected slightly more predictors than the original models, this did not lead to a large change in the top six most relevant predicators for both pain and poor function at 3 years.

The definition of OA in this target population was supported by recommendations from a group of clinicians and OA researchers in a consensus meeting. This covered adults aged 50 years or more with joint pain in at least one of several joint sites (hand, hip, knee or foot) at baseline for duration of three months or more in the previous year. Even though the participants' conditions were not confirmed by a GP diagnosis or by X-ray, the definition is in line with the clinical definition of OA suggested by most OA professional organizations such as NICE [*NICE 2008*], OARSI [*Zhang et al 2008*] and EULAR [*Pendleton et al 2000*] and as such will capture people with symptomatic OA. Considering several joints is important because most people with OA have multiple pains and may consult for say knee pain at one time but for another joint at another time. Furthermore, many interventions including the core treatments proposed by NICE and EULAR (advice and education simple analgesia, exercise and weight loss) are relevant regardless of the location of the pain. Finally, focussing on a single joint may miss people with other pain problems which are equally severe and have similar impact and hence lead to the underestimation of the prevalence of the disease.

The outcome measures (severe pain and functional limitation) used in this study have been recommended as core outcome measures for OA [*Bellamy et al 1997*] and were discussed with the same group who defined the target population of this study. The fact that the tools

used to develop these measures are validated and widely used makes them appropriate and clinically relevant to OA. Baseline levels of the outcome measures are important in prediction models for people with chronic pain conditions; hence these were included in the analysis, given that future pain is best predicted by current pain [*Dawson et al 2005*]. Different cut-off points can be used when classifying patients into different pain and functional limitations categories depending on the objective(s) of a study. In this study, even though the cut-off point for defining severe pain or functional limitation may be arbitrary, a sensitivity analysis (results not shown) was conducted in which a more extreme cut-off point (lower tertile 30% instead of the median) of functional limitation was used. The results in terms of composition of the prediction model did not change except that female sex was selected in the model. A different cut-off point of severe pain was not assessed given that a validated cut-off point [*Zelman et al 2003*] was used to derive its categories. However, changing the cut-off points of the outcome variables may lead to differences in the content and performances of the model [*Schellingerhout et al 2009*], although this is unlikely to lead to major changes in for instance the top six most important predictors.

Dichotomizing a variable is widely used and may have several advantages [*Royston et al 2006; Altman and Royston 2006; Farrington and Loeber 2000; Greenland 1995*]. For instance, dichotomizing a predictor variable facilitates simplicity of presentation of results and easy understanding of findings. Also, dichotomization is useful for practical reasons as it makes it easier for clinicians to use prognostic information when predictors are assessed on a dichotomous scale. Because of the arguments in favour of dichotomization, in this study, all the continuous variables were dichotomized (mostly using the median as cut-off point). However, dichotomizing continuous variables also has well known

disadvantages and has been criticized by several authors [*Royston et al 2006; Austin and Brunner 2004; Irwin and McClelland 2003; MacCallum et al 2002; Zhao and Kolonel 1992*]. The first disadvantage of dichotomization is loss of information - i.e. the statistical power to detect a relationship between a predictor and an outcome is reduced [*Altman and Royston 2006*].

It has been reported that dichotomizing a variable at the median is equivalent to losing one-third of that data [*MacCallum et al 2002*] and also increases the risk of a positive result being a false positive [*Austin and Brunner 2004*]. Secondly, the extent of variation in outcome between groups may be underestimated. This is where people close but on opposite sides of a cut-off point may be considered as very different rather than being similar. Thirdly, the use of two categories may conceal any non-linear relationship between the predictor and the outcome. This may also cause different predictors to remain in a final model and loss of model performance when a backward selection technique is applied in logistic regression analysis [*Schellingerhout et al 2009*].

However, a few variables have got recognized cut-off points – an example is BMI for which >30kg/m^2 is a generally accepted definition for obesity. For some variables cut-off points used in previous studies were adopted. Where this was not available the median was used, though this implies that different cut-off points may have been used in different studies, which does not facilitate comparisons between studies.

Approximately 150 variables were included during the derivation of the models. Many of these variables are related entities and hence high correlation between some variables is inevitable. Even though co-linearity is not generally regarded as a key issue in prediction modelling, if it exists between two variables it causes one to have a very strong association

with outcome and renders the other variable to have a weak or no relationship with outcome. This is usually dealt with by removing one variable from the analysis based on either the objective of the study, its importance according to the literature or if its variance inflation factor value (VIF) is greater than 5 or 10. The VIF measures the severity of multi-collinearity in a regression analysis by providing an index that quantifies how much the variance of an estimated coefficient is increased as a result of collinearity [*Mason and Perreault 1991*]. For example, there were 50 manikin variables, indicating whether there is pain present or not in a particular region. During the analysis process, some of these turned out to be associated with a lower risk of poor OA outcomes at three years. The reverse association obtained for the few manikin variables may be because those joint are not commonly affected by OA (e.g. elbow and shoulder) and the majority of the participants did not have pain at those joint sites. Even though a test of the correlation among the variables showed weak relationships between the variables, perhaps, a better approach may have been to group all manikin variables and include a single variable in the model, indicating the number of pain sites reported by individuals.

Poisson regression technique was selected as the primary method of analysis in this study as it estimates RR which is the true risk of having an outcome and the natural measure used in the construction of epidemiological indicators (PAR and NNT) assessed in this study. Logistic regression produces ORs which may not be a good approximation of RR when the outcome measure is common, as was the case in this study. However, the results of the two techniques were comparable in terms of their effect estimates (i.e. ORs and RRs), epidemiological measures (PARs and NNTs), composition and performance of the models.

Stepwise variable selection is a popular and easy way of developing a multivariable prediction model [*Steyerberg et al 2000 and 1999*]. Backward selection technique was adopted during the fitting of the models in this study as it is capable of selecting a more stable combination of predictors for a model and hence is generally preferred to a forward selection technique [*Steyerberg 2009; Schellingerhout et al 2009*]. However, neither of these stepwise procedures guarantees that the most useful model will be obtained. The techniques drop or add one predictor at a time, possibly missing a good candidate predictor and hence may fail to build an 'optimal model' [*Miller 1984*]. The procedure can also result in biased estimates of the regression coefficients due to multiple testing [*Steyerberg et al 2000*] which can sometimes lead to unproven factors to show conflicting relationships with an outcome variable. Furthermore, it results in a single final model rather than a list of good candidate models that can be compared and chosen from. The all possible subset method of variable selection would have been ideal to use, where the best model could be chosen based on their AIC or BIC values, but because of the large number of variables considered in this study it was not practicable to use as it is computationally intensive to run.

The estimates of the goodness of fit tests of the models (for both severe pain and functional limitation) showed that both statistical techniques (Poisson and logistic) fit the data well with the logistic regression analysis showing slightly better fit (based on their AIC and BIC values) compared to the Poisson regression analysis. This may be explained by the fact that the outcome measures assumed counts of zeros and ones and the logistic regression method may have fitted such data better compared to the Poisson regression method. Even though the AIC and BIC criteria were used in this study to compare the two statistical methods, they are most appropriately used to select models derived using a

similar (single) statistical method and hence the way they have been used in this study may not be entirely suitable.

Also, the performance estimates (c-statistics) of both the Poisson and logistic regression models were good which indicated that both statistical techniques predicted both outcomes considerably well with both techniques estimating slightly lower c-statistics for severe pain (0.78 and 0.79) compared to functional limitation outcome (0.88 and 0.88) respectively in the final multivariable models. Although the performance of the models was good in that they discriminated well between people with or without the outcome of interest in the derivation sample, it will require external validation of this models in a prospective sample of OA consulters in order for the model to be applicable in clinical practice as the sample used in this study mainly comprised of non-consulters. The C-statistics measure was chosen to assess the predictive performance of the models in this study because it is commonly used for binary outcomes and also reasonably easy to interpret. For instance, a C-statistic estimate of 0.7 or more is generally considered to be good as such the measure has been used in previous OA studies [*Zhang et al 2011; Yusuf et al 2011*]. However, there is no consensus cut-off value for good discrimination, similar to other predictive performance measures such as r-squared and Brier scores. Furthermore, the C-statistic is dependent on the cut-off point of the outcome measure and hence changing this cut-point may lead to different interpretation of the performance of a prediction model.

Bootstrap samples were used to internally validate the models. This technique has been found to produce accurate estimates of model performance [*Steyerberg et al 2003*]. The performance of a model tends to be more precise when higher numbers of bootstrap

replicate samples are used. A few studies have recommended that bootstrap samples of 100 is enough [*Efron and Tibshirani 1993a; Chatfield 1995*] irrespective of sample size to produce a reliable estimate, and that 200 bootstrap samples produces a more stable estimate while a plateau is reached when 500 (as chosen in this study) or more bootstrap samples are used [*Steyerberg et al 2003*]. Efron and Tibshirani [1993b] used a comprehensive theory to support bootstrap as a universal technique for internal validation. Additionally, Steyerberg et al [2001a] have also shown that bootstrapping is superior to other approaches such as split-sample and cross-validation methods in estimating internal validation. In this study, the performance estimates based on the bootstrap models did not differ much from the apparent performance of the final prediction models (i.e. optimism estimates were small). This indicates that the models had good internal validity and may be generalized to patients with similar characteristics as those in this study's target population. External validation is more important to assess the performance of these models in other settings and populations but this was not feasible within the time frame of this project and hence may be part of future work.

Few studies have attempted to estimate the contribution of predictors of poor outcome of OA in terms of adjusted PAR [*Jinks et al 2006*]. The benefits of these indicators are that they are useful in the presentation of results and may play a valuable role in risk communication to healthcare providers, administrators and the public [*Altman and Deeks 2000*]. PAR and NNT can help identify the most relevant predictors and facilitate identification of individuals at increased risk of poor long-term outcomes who may benefit from treatment. They may therefore serve as useful indicators when planning resources for the management of OA. However, the shortcomings of the measures of PARs and NNTs are that they may be difficult to interpret [*Hildebrandt et al 2006*] and therefore must be

interpreted with caution. As prediction models were developed in this study and not explanatory models, the predictors identified may not necessarily be causally associated with outcome. Furthermore, the interpretation of the PAR as maximum achievable health gains also assumes the availability of perfectly effective treatments or implies that treatments can modify all predictors which may not be a realistic assumption for many prognostic factors.

Although adjusted NNTs with 95% CIs can be calculated, their CIs calculation is computationally intensive and was therefore not done in this thesis; instead unadjusted NNTs with 95% CIs were calculated for the predictors with increased risk of poor outcome in this study. However, both adjusted and unadjusted PARs with their 95% CIs were calculated for each increased risk predictor. Although their point estimates were slightly different, their CIs did overlap which suggested that they were largely similar hence unadjusted PARs were used during the selection of the six most important predictors, also for the sake of consistency as only unadjusted NNTs were estimated for the predictors.

The ultimate aim of this study was to identify a small set of strong and relevant predictors of long-term OA outcomes. In selecting this small set of relevant predictors for severe pain and functional limitation at three years, the selection rule used by Smit et al [2006] was adopted. This rule has previously been applied to select subgroups of people at increased risk of developing anxiety in later life [*Smit et al 2007*]. This rule can help select predictors for which the highest possible health benefit (IRR and PAR) and the lowest possible effort and cost (NNT) can be achieved, given the (albeit unrealistic) assumption that interventions are completely successful. The final set of predictors for severe pain and

functional limitation was reasonably stable i.e. they showed expected associations with outcome, reflected the strongest predictors of outcome and were similar across Poisson and logistic models. As expected the reduced models showed lower predictive performance compared to the final full multivariable models, and performance for the prediction models for functional limitation was generally higher than that of the severe pain models. The predictors selected could possibly be used to in future to support the identification of people at high risk of poor outcome of OA. Future research may also investigate if targeting of some of these predictors by interventions proposed by NICE as core interventions – for example advice regarding analgesia for high levels of pain, advice regarding weight loss for obesity and advice regarding exercise for poor functional performance - can lead to improvements in long-term patient outcomes. Most of these factors (e.g. SF-36, WOMAC pain, etc) can be obtained via validated and widely used questionnaires, or can easily be measured in primary care by single item questions. Some of the identified predictors (e.g. no extended education, retirement) are non-modifiable. Although these factors may not be useful as treatment targets, they could still be helpful to identify high risk subgroups.

5.4 Conclusions

The findings of this study suggest that baseline knee pain, baseline poor physical function, low educational level, having hand pain at baseline and high BMI are the most important predictors of poor outcome in people with joint pain and OA. The predictors can potentially be used to identify subgroups of high risk patients who may be targeted by more timely or more intensive treatment in primary care.

Part 2A

Chapter Six

Evidence synthesis and meta-analysis: estimating the effects of primary care interventions for OA – Introduction and methods

This part (2A) of the thesis involved an evidence synthesis and meta-analysis presented in this and the next chapter. The main purpose of this study is to summarise available evidence on the effectiveness of primary care interventions recommended by NICE for OA (i.e. advice and information, paracetamol, topical NSAIDs and exercise) for one or more joint sites (i.e. hand, hip knee and foot) and calculate estimates of their effects on pain and functional limitation in primary care patients with OA. Although interventions to lose weight for obese patients is one of the core interventions suggested by NICE, they were not considered in this study as the study is not focusing on only obese patients. A further aim is to subsequently use these effect estimates to populate the decision model developed and presented in chapters 8 and 9 to assess the cost effectiveness of optimal primary care pathways, based on these four commonly used treatments in primary care.

6.1 Introduction

Systematic review (SR) is the process of gathering all the relevant studies that conforms to a list of well-defined criteria specified in advance to provide evidence with the aim to answer a particular review question. Clearly systematic procedures are followed during the gathering of evidence which helps to minimize bias in order to obtain accurate findings with the intention of making appropriate and correct decisions or conclusions about one or more well defined question(s). The process entails the collation of the relevant studies, extraction of the study results, assessment of risk of bias among the studies, summarizing effects sizes across the studies, examination of differences among the studies (heterogeneity and publication bias) and interpretation of the findings. Meta-analysis (MA) is often carried out within SRs and the technique was first used by Glass [1976]. The

technique employs statistical methods to quantitatively combine the results of the selected individual studies with the aim of providing more accurate estimates of the effects of healthcare interventions.

6.2 Objectives

The four interventions considered in this review are recommended by NICE [*NICE 2008*] and other professional organizations such as OARSI [*Zhang et al 2008*] and EULAR [*Pendleton et al 2000; Zhang et al 2007a*] as core treatment options for the management of OA as they are considered to be safe options to reduce pain and improve function. Because there are already numerous published systematic reviews and meta analyses on the interventions considered in this study, it was possible to search for evidence in an efficient way by identifying relevant trials from available SRs and conducting an updated search of individual trials only.

The specific objectives of this part of the study are;

(1) To carry out an evidence synthesis by identifying relevant randomised controlled trials (RCTs) on the effectiveness of the four interventions in primary care populations.

(2) To assess the risk of bias within each of the selected RCTs.

(3) To extract relevant data from the selected RCTs to calculate effect size estimates for the four interventions.

(4) To examine evidence of heterogeneity among effect estimates and subsequently use appropriate methods to pool them.

(5) To explore the possibility of publication bias.

6.3 Methods

This section illustrates the methods used to carry out the evidence synthesis and meta-analysis.

6.3.1 Selection criteria

Included in the evidence synthesis were randomized controlled trials (RCTs), published in English, which evaluated the effectiveness of advice/information regarding self-management approaches, paracetamol, topical NSAIDs and exercise amongst patients aged 45 years and above diagnosed (radiographically or symptomatically) with OA at one or more joint (hand, hip, knee or foot) sites.

More detailed inclusion criteria were as follows:

1. Participants must be adults aged 45 years or over and diagnosed with OA either clinically or radiographically.

2. Participants must suffer from OA or joint pain/disability at one or more joint sites and the sites must involve the hand, hip, knee or foot.

3. RCTs must report data on either pain or functional limitation. The WOMAC pain and/or functional limitation (SF-36) would be considered if more than one pain and/or functional limitation outcome measures were reported.

4. RCTs must examine the effect of advice about self-management, paracetamol, topical NSAIDs or exercise for OA.

5. RCTs must include a follow-up period of at least 3 weeks in order to include as many studies as possible in this review.

6. Participants must be primary care patients; for countries without a structure of primary care such as the UK, patients should be recruited from settings where physician act as a gate keeper, patients should have direct access to care (e.g. without patients clinics, occupational care or emergency care) or patients should be recruited from community based studies.

Exclusion criteria were as follows:

1. Participants presenting for surgery.

2. Secondary care patients i.e. patients referred by other care providers.

3. RCTs not published in English language.

4. RCTs without full report either online, in grey literature (i.e. not formally published in sources such as books or reports) or in peer-reviewed journals.

It was important that the four interventions considered were compared within consistent comparison groups to enable formal pooling of the effect estimates. The comparison groups were decided during a consensus meeting with clinicians. The comparisons of interest were as follows: (1) advice versus no treatment, (2) simple analgesics versus advice/placebo/no treatment, (3) topical NSAIDs versus advice/placebo/no treatment and (4) exercise versus advice/simple analgesics/no treatment.

6.3.2 Information sources

A full search of all healthcare databases was not feasible given the timeline of this project. Therefore instead, systematic reviews (SRs) and meta-analyses (MAs) were searched in the Cochrane register from January 1990 to August 2010. The 20 year time frame was considered because it was expected that this would cover large and most recently updated SRs/MAs in which relevant and up to date clinical studies would be identified. Subsequently, an additional search in MEDLINE, covering the period from the year 2000 to August 2010 was conducted to identify individual trials that were not yet included in reviews.

Over 2,500 new SR/MAs and many more individual articles reported in English are indexed in MEDLINE database annually and the majority are concerned with estimating effectiveness of interventions [*Moher et al 2007*], rendering this the database of choice for many researchers aiming to identify intervention studies. Also, the SRs searched have generally utilized a wider range of databases, which means that they would have identified RCTs listed in many other databases. Therefore, it was decided not to search bibliographic databases such as EMBASE and CINAHL (Cumulative Index to Nursing and Allied Health Literature) for this evidence synthesis.

In situations where there was more than one SR for the same intervention, the latest and more updated SR/MA containing most studies was used and others were cross-checked to ensure that all RCTs were included in the analysis. Reference lists from all retrieved systematic reviews and RCTs were also checked to identify additional potentially eligible RCTs.

6.3.3 Search strategy and identification of studies

The search terms were developed in consultation with a systematic reviewer based at the Arthritis Research UK Primary Care Centre. Titles, abstracts and keywords of each article were searched for the main medical subject heading (MeSH) key words: osteoarthritis, family practice, general practice, primary health care, community health services, ambulatory care, anti-inflammatory agents – non steroidal, anti-rheumatic agents and exercise therapy and each were exploded. The complete search strategy used for the Cochrane and MEDLINE databases can be found in appendix 7a and 7b respectively.

Four reviewers Jerome Wulff (JW), Danielle van der Windt (DvdW), Milisa Blagojevic (MB) and Sue Jowett (SJ) contributed to the identification of eligible studies. Eligibility assessment of each study was performed by the author of this thesis (JW) whilst one of the supervisors (DvdW) for this project double checked that the studies selected met the eligibility criteria. The two other reviewers (MB and SJ) helped to make decisions in case of uncertainty. Initially, titles and abstracts of reviews and individual RCTs were screened to eliminate duplicates and individual studies that did not meet the eligibility criteria. Full texts of the remaining reviews were examined to help identify individual RCTs which were added to those individual RCTs already selected from the Medline database.

6.3.4 Data extraction

Data were extracted from the RCTs included in this review to enable description of the characteristics of the trials, assessment of risk of bias, and calculation of effect estimates. This included information on:

(i) Country and setting: primary care, out-patient clinic, others.

(ii) Study population: gender (% female), mean age or age range (years), location of joint pain.

(iii) Study design: concealment of allocation, blinded assessment of outcome, loss-to-follow-up (%) and analysis according to intention to treat.

(iv) Interventions: type of intervention (advice, simple analgesics, topical NSAIDs exercise): duration and frequency or dose of treatment.

(v) Outcome measures (pain and functional limitation): instrument used and timing of assessment.

Absolute mean or changes in means at the end of treatment and their respective standard deviation (SD) estimates of the outcome measures were also extracted from original studies and used to calculate treatment effect estimates (i.e. standardized mean difference) where possible. If a study examined more than two treatment groups (e.g. 2 types of exercise treatments groups compared to a control group) the primary (standard) intervention group and control group are used. Both intention to treat (i.e. including all the subjects randomized to interventions in a study) and per protocol (i.e. including only subjects who did not deviate from a study's protocol) analyses were considered. If a study reported both types of analyses results from the former analysis was used. Two authors were contacted for additional data but only one responded that data was not available as the study was conducted over 20 years ago. The other author did not respond. Data extraction was performed by the author of this thesis.

6.3.4.1 Summary measure of effect estimates

Standardized mean difference (SMD) was used as the measure of intervention effect estimate since all the studies included in this review recorded their outcome data for pain and functional disability on a continuous scale. Given that different instruments and scales were used to measure pain or functional limitation, standardized mean difference (SMD) was calculated for each comparison and each study using the formula recommended by Hedges and Olkin [1985] to enable valid comparison of results: The SMD is obtained by dividing the difference in mean outcome values by the pooled standard deviation:

$$SMD = \frac{\overline{X}_T - \overline{X}_C}{\sqrt{\frac{S_T^2(n_T - 1) + S_C^2(n_C - 1)}{n_T + n_C - 2}}} \qquad (20)$$

where $\overline{X}_T$ and $\overline{X}_C$ are either the change or absolute mean outcome values of treatment and control groups reported at the end of treatment, S_T^2 and S_C^2 are their respective variances and n_T and n_C are their sample sizes. In this study, effect estimates of 0.2, 0.5 and 0.8 were described as small, moderate and large respectively as suggested by Cohen [1977]. The summary of means, standard deviations and number of participants for the studies used to pool effect estimates for the four interventions considered in this study are shown in appendices 8a to 8d.

6.3.5 Assessment of risk of bias of individual RCTs

Risk of bias of the individual RCTs included in this review was assessed to obtain a measure of quality of the publications. The Cochrane Risk of Bias tool was used [*Higgins et al 2011*] for this purpose. The tool consists of six domains, each of which is scored as either high risk, low risk or unclear. The six domains are selection bias (i.e. random sequence generation and allocation concealment), performance bias (blinding of patients/personnel), detection bias (blinding of outcome assessors), attrition bias (incomplete outcome data due to loss to follow up), selective outcome reporting bias, and other biases not covered above (e.g. fraudulent study). The full Risk of Bias Cochrane tool and the criteria for judging risk of bias can be found in appendix 9.

In the Cochrane Risk of Bias tool each domain is scored separately because the use of a scale or check list (composite scores) that numerically summarize multiple components into a single number have been found to be misleading and unhelpful [*Juni et al 2001; Egger et al 2001*]. The summated scores do not provide information on individual sources of bias and may obscure severe flaws in trial design or conduct if this covers only one domain, for example very high and selective drop-out rate.

The domains considered in this review were adequacy of randomization, concealment allocation, blinding of outcome assessors and loss to follow up and were assessed by the author of this thesis using the same three possible judgments (i.e. high risk, low risk and unclear risk) as applied in the Cochrane tool. Blinding of outcome assessors was the only form of blinding considered as it was not possible to blind patients and care providers for

interventions such as advice and exercise even though blinding of patients and clinicians was possible in the analgesics and topical NSAIDs studies.

Although the aim of this study was to include as many studies as possible to obtain the most updated effect estimates for the four interventions considered in this study, sensitivity analyses were performed omitting trials with high risk of bias in at least one of the domains considered, and this involved the analyses of effectiveness of advice and exercise interventions.

6.3.6 Meta - analysis

After extracting the information from the trials to obtain their individual effect estimates, the pooled (overall) effect estimates for each intervention were calculated using a quantitative technique known as meta-analysis. The process involves assessing the presence of heterogeneity among the studies and subsequently choosing the correct model (fixed or random effects) to pool the estimates, assessing what attributes of the studies may account for heterogeneity if present, and examination of publication bias. These steps are described in detail below.

6.3.6.1 Evaluation of heterogeneity

When similar effect estimates are observed among the studies, the studies are said to be homogeneous. On the contrary, when the effect estimates are found to vary, they are described as heterogeneous. Formally examining heterogeneity of effects across studies is very important in MA as it helps to decide what model may be most appropriate, or in case

of wide and persistent heterogeneity, not to calculate a pooled effect estimate. As outlined below, the fixed effects (FE) model assumes study estimates to be homogeneous, whereas the Random Effects (RE) model has the ability to account for some degree of heterogeneity.

A test for heterogeneity examines the null hypothesis that all studies in a MA have the same effect where a non-significant result ($p > 0.05$) is taken as evidence of homogeneity. A commonly used approach for evaluating presence of significant heterogeneity between studies is the chi-squared test (Cochran's Q-statistics) [*Cochran 1954*]. This statistic however, does not provide information on the extent of the detected heterogeneity, and it may also have low power to detect significant findings when there are few studies [*Alexander et al 1989*]. The Q statistic is estimated as $Q=\sum w_i (\Theta_i - \Theta_{iv})^2$, where $\mathbf{w_i}$ is the inverse variance estimate (weight) of study i and θ_i its effect estimate and θ_{iv} is the pooled effect estimate. In order to provide a more meaningful measure of the extent of heterogeneity, Higgins and Thompson [2002] introduced a selection of useful indices, with I^2 being the easiest to interpret and was hence used in this study.

It is given as (*Higgins and Thompson 2002*):

$$I^2 = \frac{\tau^2}{\tau^2 + \sigma^2} \qquad (21)$$

where σ^2 is the within study variance and τ^2 is the between study variance. It can be viewed as representing the percentage of the total variation of the estimated effects across studies that is due to between-study variation rather than to chance. Higgins *et al.* [2003] provide a

rough guide for its interpretation as follows; 0% to 25% - low heterogeneity, 26% to 50% - moderate heterogeneity, 51% to 75% - substantial heterogeneity and 76% to 100% - considerable heterogeneity.

Both Q statistic and I^2 index have to be interpreted with caution in case of small sample size, as was the case for studies that examined advice and information, simple analgesia and topical NSAIDs in this review.

In this study, fixed effects (FE) models were employed if assessment of heterogeneity showed that the individual studies were homogeneous, i.e. if the I^2 estimate of a model was $\leq$ 50% – otherwise random effects (RE) models were used. If substantial or more heterogeneity was detected, study characteristics such as publication year, quality of study, differences in interventions, outcome measures, study design, or joint affected, would have been assessed as possible sources of heterogeneity using meta-regression analysis.

The section below illustrates in detail FE and RE models.

6.3.6.2 Fixed and random effect models

The two main methods used for combining data of several studies to obtain overall effect size estimates are fixed effects (FE) and random effects (RE) models. The FE model assumes that the observed variation of the effect estimates of individual studies is entirely due to sampling variation (chance) - that is, the true effect estimate is the same in each study. The approach therefore assumes effect estimates to be homogeneous and only

considers within study variability. Inverse variance method is most commonly used though other measures such as the standard error, sample size, etc, of a study can also be used as weights.

The inverse variance is given by;

$$w_i = \frac{1}{\sigma_i^2} \qquad (22)$$

where σ_i^2 is the variance within study *i*. The method will give more weight to larger studies which have small variance estimates than smaller studies which have larger variance estimates and thus minimises the variability of the pooled effect estimate.

The RE model assumes effect estimates to be heterogeneous and considers both within and between study variability. It assumes that individual studies do not share a common effect size estimate, but rather the variation of the effects across studies follows, most commonly, a normal distribution whose mean equals the true overall effect estimate [*DerSimonian and Laird 1986*].

Developed by Der Simonian and Laird (1986), it is a variation of the inverse variance method of FE model, with weights given by:

$$w_i = \frac{1}{\sigma_i^2 + \tau^2} \qquad (23)$$

where τ^2 is the between study variance calculated as:

$$\tau^2 = \frac{Q-(k-1)}{\sum w_i - \left(\frac{\sum w_i^2}{\sum w_i}\right)} \qquad (24)$$

where Q is the chi-square test of heterogeneity statistic (explained above), *k* is the number of contributing studies and w_i is the individual study weight as defined in equation 23.

The pooled effect estimate is then formulated as:

$$\theta_{iv} = \frac{\sum w_i \theta_i}{\sum w_i} \qquad (25)$$

where, θ_i is the effect estimate of an individual study and w_i is the inverse variance weights defined for FE and RE models in equations (22) and (23) respectively specified above.

RE models offer more weight to the results of smaller studies than the FE models which may be unfavorable as small studies often lack quality and may be more prone to publication bias [*Liberati et al 2009*]. In other words, the weights assigned to studies in RE models are much more similar to each other than in FE models. Hence, if the pooled estimate of a RE model differs from that of FE model for the same intervention, it implies that the average estimate from smaller studies differs from the average of larger studies. The confidence intervals for the pooled effect estimate in RE models are wider with correspondingly larger p-values compared to FE models because within and between study variances are combined. The greater the value of between study variance τ^2 the greater the difference between the RE and FE weights. Obviously if τ^2 is very small RE and FE estimates will be very similar.

In order to decide which model may be more appropriate, it is important to evaluate presence and extent of heterogeneity formally, which has been explained above. In this study, FE models were employed since assessment of heterogeneity revealed that the individual studies considered may be assumed to be homogeneous since the I^2 estimate of the models were $\leq$ 50%. RE models were carried out (data not shown) but their results were the same as that of the FE models.

6.3.6.3 Evaluation of publication bias

Publication bias occurs when studies with significant findings (or larger / more positive effect estimates) are more likely to be published than those failing to show such significant or positive findings. Other sources of bias include language bias (i.e. where studies published only in a particular language, for example English, are selected and included in a review), availability bias (i.e. where studies easily accessible to the researcher are selected and included in a review), cost bias (i.e. where studies that are available free of charge or at low cost are selected and included in a review), familiarity bias (i.e. where studies that from one's own area of specialty/discipline are selected and included in a review) and outcome bias (i.e. where studies of some primary outcome measure of interest to a researcher but not other outcomes are selected depending on the direction and statistical significance of the results are included in a review) [*Rothstein et al 2005*]. These biases generally cause the overall effect estimate of an intervention to be over-estimated.

Funnel plots are usually used to assess the risk of publication bias and Egger's test [*Egger et al 1997*] is used to confirm if this risk of publication bias is plausible. Funnel plots are essentially simple scatter plots of treatment effect estimates (typically on the x-axis) from individual studies against either a measure of each study's sample size or precision.

In this review, treatment effects were plotted against the SE of the treatment effects. The SEs are plotted on a reverse scale (i.e. zero at the top with increasing values to the bottom on the vertical axis) resulting in the effect estimates from small studies being scattered widely at the bottom of the graph with the spread narrowing towards the top of the graph for larger studies. The use of SEs also facilitates the plotting of a triangular region on the funnel plot graph within which 95% of the studies are expected to lie if both bias and heterogeneity are absent. That is, in the absence of publication bias a typical funnel plot would look like a symmetrical inverted funnel. On the other hand, an asymmetrical appearance of the funnel plot with a gap at the bottom right corner of a graph would indicate presence of bias which is caused by non-availability of published studies that failed to show significant results. Sterne et al [2000] warned that funnel plot asymmetry should not always be linked with publication bias, but rather as means of displaying and evaluating small study effects as they have the tendency of reporting large effects estimates due to poor methodological quality compared to those estimated in larger studies.

Egger's [1997] test is usually used to formally examine whether the association between estimated treatment effect and some measure of study size (e.g. SE of the intervention effect) is greater than might be expected to occur by chance. The test performs a linear regression of the intervention effect estimates on standard errors, weighted by the inverse of the variance of the intervention effect estimate to ensure that the regression estimates are

not dominated by smaller studies. This looks for a straight line relationship between intervention effects and standard errors under the null hypothesis of no small study effects.

All analyses were performed using STATA version 11 [*StataCorp. 2009*].

Chapter Seven

Evidence synthesis and Meta-analysis: estimating the effects of primary care interventions for OA – Results, discussions and conclusion

7.1 Selected studies

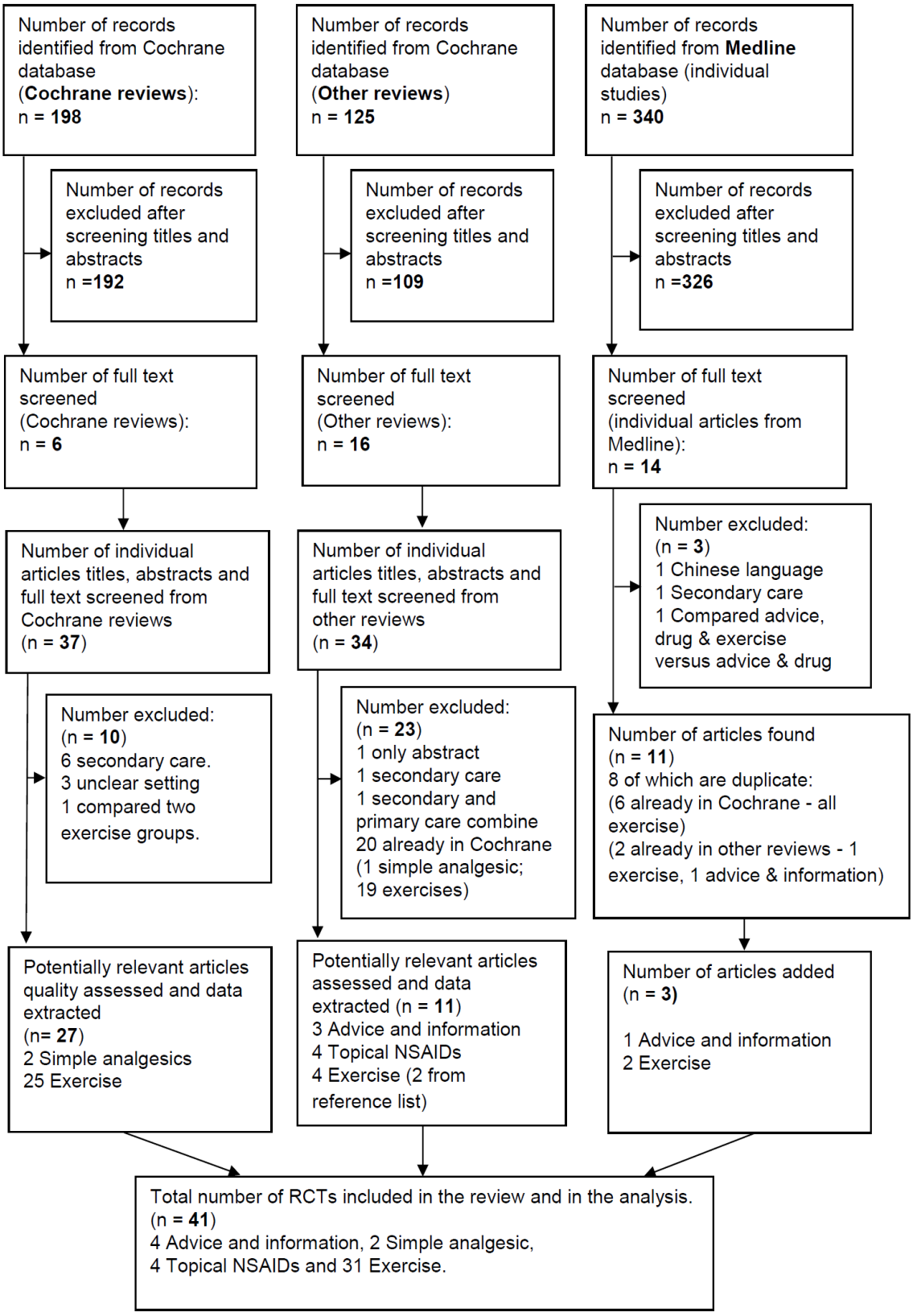
The total number of hits obtained after running the search in Cochrane database was 2505. Of these, 198 were Cochrane reviews, 125 were other reviews and the remaining 2182 were health technology assessments and economic evaluation studies. After screening titles and abstracts and adjusting for duplicates, 192 Cochrane reviews and 109 other reviews were excluded because they did not meet the inclusion criteria, leaving 6 Cochrane reviews and 16 other reviews to be screened individually. Full texts were screened and 37 and 34 individual RCTs were identified from Cochrane and other reviews respectively. The full text of each of these individual RCTs was then examined in detail, resulting in exclusion of 10 RCTs from the Cochrane reviews (6 secondary care, 3 unclear settings and 1 compared two exercise interventions) and exclusion of 23 from the other reviews (1 only abstract found, 1 secondary care, 1 secondary and primary care patients combined and 20 RCTs were already identified from the Cochrane reviews). As a result, the total number of eligible RCTs subjected to quality assessment and data extraction was 27 (2 investigating simple analgesics and 25 exercise interventions) from Cochrane reviews and 11 (3 investigating advice and information, 4 topical NSAIDs and 4 exercise interventions – 2 from reference list) from other reviews.

Additionally searching the MEDLINE database for individual RCTs resulted in 340 hits. After screening the titles and abstracts and removing duplicates, 326 articles were excluded because they did not meet the inclusion criteria, leaving 14 individual RCTs to be examined further. 11 remained after further excluding 3 articles upon complete reading of full text (1 Chinese language, 1 secondary care and 1 compared three interventions combined versus two interventions combined). Furthermore, 6 were already identified

from the Cochrane reviews and 2 already from other reviews leaving only 3 additional RCTs from the MEDLINE search.

In summary, a total of 41 RCTs (27 from Cochrane reviews, 11 from other reviews and 3 from the Medline search) were included in this review of which 4 examined advice and information [*Keefe et al 1990; Heuts et al 2005; Wetzels et al 2008; Ravaud et al 2009*], 2 examined simple analgesia [*Case et al 2003; Pincus et al 2004*], 4 examined topical NSAIDs [*Grace et al 1999; Bookman et al 2004; Roth and Shainhouse 2004; Niethard et al 2005*] and 31 examined exercise interventions (see table 7.4). The two RCTs identified from reference checking investigated exercise [*Wang et al 2007; Hinman et al 2007*]. Figure 7.1 below illustrates the literature search and results of study selection for this review.

Figure7.1 Flow chart presenting results of literature searches and study selection



7.2 Study characteristics

The main characteristics of studies included in the review are presented in tables 7.1 to 7.4, and a summary is given below.

7.2.1 Study settings

Although the review originally aimed to identify RCTs carried out in a primary care setting, the setting was not clear for a few studies, particularly those investigating use of simple analgesia and topical NSAIDs. It was therefore decided to include those RCTs in which participants seemed to have been recruited from the community (population) or from setting where people have direct access to medical care. Most of these RCTs were carried out in the USA where there is no similar primary care setting as in the UK or some other European countries (see tables 7.1 to 7.4).

7.2.2 Number of study participants

The 41 studies included in this reviews involved 6,715 subjects assessed for pain and 5322 subjects assessed for functional disability. The numbers of participants used to calculate the effects estimate of the four interventions are summarized in their respective results sections (7.4.1 to 7.4.4). There was marked variation between the 41 individual studies included in this review in their number of subjects recruited. The mean number of subjects recruited to RCTs investigating advice and information intervention was 193 (range 67 to

327), for RCTs on simple analgesia 200 (range 57 to 343), for RCTs investigating topical NSAIDs 198 (range 68 to 322), and for exercise trials 156 (range 30 to 783).

7.2.3 Description of interventions

Three out of four of the advice and information studies used usual care as their control group whereas one used an educational leaflet as control (table 7.1). The two analgesia studies compared paracetamol (1000 mg 4 times daily) with placebo (table 7.2) whereas all the four topical NSAIDs studies compared topical diclofenac with placebo (table 7.3). A wide range of therapeutic exercise routines were assessed among the 31 exercise intervention studies (table 7.4). These comprised of exercise programs delivered individually to the patient, delivered to a small group of people in a class-base format and exercise routines mostly undertaken by the patient at home. The treatment content varied from quadriceps muscle strengthening involving leg raising exercises only and aerobic walking routines [*Ettinger et al 1997; Talbot et al 2003; Messier et al 2004*] to comprehensive programmes including lower limb strengthening, upper limp and truncal muscle strengthening and balance coordination [*van Baar et al 1998b; Peloquin et al 1999; Deyle et al 2000; Bennell et al 2005*]. The control groups compared with the various exercise programs included waiting list, education classes, usual care/activities and no intervention.

7.2.4 Joint affected

The knee was the most commonly affected joint among the studies considered in this review. Two advice and information studies involved participants with knee OA only,

whilst the other two involved subjects with knee or hip OA. One simple analgesia study focused on participants with only their knee OA whilst the other study involved subjects with either their knee or hip affected. All of the four topical NSAIDs studies involved subjects with only their knees affected. For the 31 studies which considered exercise intervention, 17 studies (55%) had only their knees affected, 12 studies (39%) had their knees or hips affected whilst only 2 studies (6%) had their hips affected. The few trials which investigated subjects with hand or foot OA did not meet the inclusion criteria and none assessed treatment in people with OA regardless of the joint affected.

7.2.5 Outcome assessment and duration of treatment

Four exercise studies [*Belza et al 2002; Talbot et al 2003; Keefe et al 2004 and Ravaud et al 2004*] and one simple analgesia study [*Pincus et al 2004*] reported on pain only whilst the remaining studies reported on both pain and function. In this review, outcome values at the end of treatment were extracted for all the studies and used to calculate the effect estimates.

The most common duration of treatment for the interventions was 6 months for advice and information (longest duration was 6 months), 3 weeks for topical NSAIDs (longest duration is 3 months), 12 weeks for exercise (longest duration is 2 years) and the longest duration for simple analgesia was 3 months. The exercise studies reported the longest treatment duration (i.e. 3 years).

The duration of treatment of the studies varied. The mean duration for the advice and information studies was 19 weeks ranging from 12 weeks to 24 weeks, that for the simple

analgesia studies was 9 weeks (range 6 to 12 weeks), for topical NSAIDs studies it was 5.5 weeks (range 3 to 12 weeks) and for the exercise studies it was 19 weeks (range 4 to 96 weeks).

To calculate the effect estimates of the interventions, the measures of pain and function were recorded either as mean change in score from baseline to end of treatment or the final outcome score at the end of a treatment period. This does not pose a problem in meta-analysis of mean differences since both scores are considered to be addressing the same underlying intervention effect in RCTs [*Deeks et al 2011*]. Also, on average, differences in mean final scores will be the same as difference in mean change score, if randomization has been successful and baseline values of outcome measures are similar. One important advantage of using change score is that, it ensures that differences in baseline (pre-treatment) scores between groups do not affect the estimates of effect size. The various tools used to measure the outcome variables (pain and functional limitation) in each study are shown in tables 7.1 to 7.4 below. For outcome measurement tools such as SF-36 where high scores imply better/improved function, a negative sign is applied to such scores to change the direction of effect to match tools such as WOMAC where low scores imply reduced pain or improved function.

The study specific data on the number of patients and the mean scores (and corresponding standard deviations) of outcome measures extracted which were used to pool the effect estimates for each of the four interventions (advice and information, simple analgesia, topical NSAIDs and exercise) are shown in appendices 8a to 8d respectively.

Table 7.1 Characteristics of RCTs investigating effectiveness of advice and information

Author, Year (Country)	Gender (%) Female	Age (Years)	Setting	Description of interventions (Advice and Information)	Treatment duration	Outcome	Joint affected (OA)
Keefe, 1990 (USA)	NS	64 mean	Primary care patients; Diagnosis: clinical and X-ray	Arthritis Education: n= 36 Standard/usual Care: n=31	6 months	Pain (AIMS), Function (AIMS)	Knee
Heuts, 2005 (Holland)	60	52 mean	Primary care patients; Diagnosis: clinical and X-ray	Advice on self management: n = 132 Usual Care: n= 141	3 months, 21 months follow up	Pain (VAS), Function (SF-36)	Knee and hip
Wetzels, 2008 (Holland)	76	75 mean	Primary care patients; Diagnosis: clinical	Nurse-based self-management program: n=51 Control group (educational leaflet) : n= 53	6 months, 12 months follow up	Pain (AIMS), Function (AIMS)	Knee and hip
Ravaud, 2009 (France)	75	64 mean	Primary care patients; Diagnosis: clinical and X-ray	Education on self management (Standardized Consultation): n = 146 Usual Care: n= 181	4 months	Pain (VAS), Function (WOMAC)	Knee

NS – Not stated

Table7.2 Characteristics of RCTs investigating effectiveness of simple analgesia

Author, Year (Country)	Gender (%) Female	Age (Years)	Setting	Description of interventions (Paracetamol)	Treatment duration	Outcome	Joint affected (OA)
Case, 2003 (USA)	56	62 mean	Primary care patients; Diagnosis: clinical and X-ray	Paracetamol (1000 mg 4times daily) n=29 Placebo: n=28	12 weeks	Pain (WOMAC) Function (WOMAC)	Knee
Pincus, 2004 (USA)	62	64 mean	Setting not stated ; Diagnosis: clinical and X-ray	Paracetamol (1000 mg 4times daily) n= 171 Placebo: n=172	6 weeks	Pain (VAS)	Knee or Hip

Table7.3 Characteristics of RCTs investigating effectiveness of topical NSAIDs

Author, Year (Country)	**Gender (%) Female**	**Age (Years)**	**Setting**	**Description of interventions** (Topical NSAIDs)	**Treatment duration**	**Outcome**	**Joint affected (OA)**
Grace, 1999 (Canada)	61	62 mean	Setting not stated; Diagnosis: clinical and X-ray	Topical Diclofenac: n=34 Placebo: n=34	3 weeks	Pain (WOMAC) Function (WOMAC)	Knee
Bookman, 2004 (Canada)	63	61.8 mean	Primary care patients; Diagnosis: clinical and X-ray	Topical Diclofenac: n=84 Placebo: n=79	4 weeks	Pain (WOMAC) Function (WOMAC)	Knee
Roth, 2004 (USA)	68	64 mean	Primary care patients; Diagnosis: clinical	Topical Diclofenac: n=163 Placebo: n=159	12 weeks	Pain (WOMAC) Function (WOMAC)	Knee
Niethard, 2005 (Germany)	64	66 mean	Primary care patients; Diagnosis: clinical and X-ray	Topical Diclofenac: n=117 Placebo: n=120	3 weeks	Pain (WOMAC) Function (WOMAC)	Knee

Table7.4 Characteristics of RCTs investigating effectiveness of exercise

Author, Year (Country)	Gender (%) Female	Age (Years)	Setting	Description of interventions (Exercise)	Treatment duration	Outcome	Joint affected (OA)
Bautch, 1997 (USA)	73	69 mean	Primary care and community based patients; Diagnosis: clinical and X-ray	Individual program: 36 sessions ROM/ walking and education classes. n=15 Education Control: Education classes. n=15	12 weeks	Pain (VAS) Function (AIMS)	Knee
Ettinger, 1997 (USA)	69	69 mean	Community based patients; Diagnosis: clinical and X-ray	Class based program. 36 sessions of aerobic walking: n=144 Education control: Education classes plus monthly telephone calls. n=75	12 weeks.	Pain (FAST) Function (FAST)	Knee
van Baar, 1998 (Holland)	79	68 mean	Primary care patients: Diagnosis: clinical and X-ray	Individual program: 17 sessions of physiotherapy + GP education. n=54 Control group: GP education. n=59	12 weeks	Pain (VAS) Function (IRGL)	Knee or hip
Maurer, 1999 (USA)	42	64 mean	Outpatient clinic patients; Diagnosis: clinical and X-ray	Individual program: 24 sessions of unilateral quadriceps strengthening only. n=49 Control group: 4 Education classes. n=49	8 weeks	Pain (WOMAC) Function (WOMAC)	Knee
O'Reilly, 1999 (UK)	66	62 mean	Primary care and community based patients; Diagnosis: clinical and X-ray	Home program: Quads/hamstring strengthening, lifestyle advice plus 4 home visits. n=108 Control group: Lifestyle advice. n=72	6 months	Pain (WOMAC) Function (WOMAC)	Knee

Table7.4 continued

Author, Year (Country)	Gender (%) Female	Age (Year)	Setting	Description of interventions (Exercise)	Treatment duration	Outcome	Joint affected (OA)
Peloquin, 1999 (Canada)	70	66 mean	Community based patients; Diagnosis: clinical	Class-based program: 36 sessions of aerobic and strength/stretching exercise. n=59 Control group: Usual daily activity plus 12 education classes. n=65	12 weeks	Pain (AIMS) Function (AIMS)	Knee
Deyle, 2000 (USA)	60	61 mean	Community based patients: Diagnosis: clinical	Individual program: 8 sessions of manual therapy/strengthening n=33 Control group: Ultrasound (sub-therapeutic): n=36	4 weeks, 8 weeks follow up	Pain (WOMAC) Function (WOMAC)	Knee
Hopman-Rock 2000 (Holland)	80	65 mean	Community based patients: Diagnosis: clinical And X-ray.	Class-based program: Education and exercise. n=45 Control group: Waiting list. n=37	6 weeks	Pain (VAS) Function (IRGL)	Knee or hip
Patrick, 2001 (USA)	86	66 mean	Community based patients; Diagnosis: clinical	Range-of-motion, maintenance of muscle strength. n=125 Control group: Usual activities. n=124	20-weeks	Pain (HAQ) Function (HAQ)	Knee or hip
Baker, 2001 (USA)	74	69 mean	Community based patients ; Diagnosis: clinical and X-ray.	Home program: Muscle strengthening plus 12 visits. n=23 Nutrition education control: 7 home visits n= 23	16 weeks	Pain (WOMAC) Function (WOMAC)	Knee

Table7.4 continued

Author, Year (Country)	Gender (%) Female	Age (Years)	Setting	Description of interventions (Exercise)	Treatment duration	Outcome	Joint affected (OA)
Fransen, 2001 (Australia)	70	66 mean	Community based and outpatients ; Diagnosis: clinical and X-ray.	Individual or class-based: 16 sessions of muscle strengthening and aerobic exercise. n=83 Control group: Waiting list. n=43	8 weeks	Pain (WOMAC) Function (WOMAC)	Knee
Halbert, 2001 (Australia)	59	69	Primary care and community based patients; Diagnosis: clinical	Individual program: 3 sessions per week of physical activity plus advice. n= 37. Control group: Nutrition pamphlet. n=32	12months	Pain (WOMAC) Function (WOMAC)	Knee or hip
Thomas, 2002 (UK)	65	62 mean	Primary care and community based patients; Diagnosis: clinical	Home program: Muscle strength training, bilateral with Theraband. n=467 Control group: Monthly telephone call. n= 316	24 months	Pain (WOMAC) Function (WOMAC)	Knee
Topp, 2002 (USA)	72	63 mean	Community based patients ; Diagnosis: clinical	Class-based program: 16 sessions of muscle strengthening with Thera-band elastic bands. n=67 Control group: No intervention. n=35	16 weeks	Pain (WOMAC) Function (WOMAC)	Knee
Belza, 2002 (USA)	86	66 mean	Community based patients ; Diagnosis: clinical and X-ray.	Group program: Exercise (water): n=125 Control (waiting list): n= 124	5 months	Pain (VAS)	Knee or hip

Table7.4 continued

Author, Year (Country)	Gender (%) Female	Age (Years)	Setting	Description of interventions (Exercise)	Treatment duration	Outcome	Joint affected (OA)
Foley, 2003 (Australia)	49.5	71 mean	Community based patients ; Diagnosis: clinical	Aquatic exercise: Stretching and strengthening exercise. n=35 Land-based (Gym): n=35 Strengthening exercise Control group: 3 (2 weekly) telephone calls. n=35	6 weeks.	Pain (WOMAC) Function (WOMAC)	Knee or hip
Quilty, 2003 (UK)	NS	67 mean	Community based patients ; Diagnosis: clinical and X-ray.	Class based program: Individual exercise plus 9 physiotherapy sessions. n=43 Control group: No co-intervention. n=44	5 months:	Pain (VAS) Function (WOMAC)	Knee
Talbot, 2003 (USA)	77	70 mean	Community based patients ; Diagnosis: clinical and X-ray.	Home program: 12 ASMP classes plus home-based pedometer walking program. n= 17 Control group: Arthritis Self-Management Program (ASMP) classes. n=17	12 weeks, 24 weeks follow up	Pain (McGill)	Knee
Hughes, 2004 (USA)	83	74 mean	Community based patients ; Diagnosis: clinical and X-ray.	Class-based program: 24 sessions of muscle strengthening plus aerobic walking plus education. n=68 Control group: Arthritis Help book. n=43	8 weeks,	Pain (WOMAC) Function (WOMAC)	Knee or hip

Table7.4 continued

Author, Year (Country)	Gender (%) Female	Age (Years)	Setting	Description of interventions (Exercise)	Treatment duration	Outcome	Joint affected (OA)
Keefe, 2004 (USA)	50	59 mean	Community based and outpatients; Diagnosis: clinical and X-ray.	Class-based program: 36 aerobic sessions and strengthening sessions. n=16 Control group: Usual care. n=18	12 weeks	Pain (AIMS)	Knee
Messier, 2004 (USA)	70	69 mean	Community based patients; Diagnosis: clinical and X-ray.	Class-based program: 48 sessions of strengthening and aerobic walking plus telephone n=80 Control group: Healthy lifestyle, 3 monthly education plus 8 telephone calls. n=78	6 months, 18 months Follow up	Pain (WOMAC) Function (WOMAC)	Knee
Ravaud, 2004 (France)	68	66	Primary care patients; Diagnosis: clinical and X-ray	Land-based exercise: Via video tape: 4 weekly 30 minute sessions. n= 352 Control group: Usual care. n=388	24 weeks	Pain (VAS)	Hip
Lin, 2004 (UK)	88	69 mean	Community based patients; Diagnosis: clinical	Group program: Water exercise. n=66 Control group: Education. n= 40	12 months	Pain (WOMAC) Function (WOMAC)	Knee or hip
Cochrane, 2005 (UK)	63	70 mean	Primary care and community based patients; Diagnosis: clinical	Aquatic exercise: Stretching, strengthening and aerobic exercises. n=153 Control: n=159 Telephone calls	3 months, 9 months follow up	Pain (WOMAC) Function (WOMAC)	Knee or hip

Table7.4 continued

Author, Year (Country)	Gender (%) Female	Age (Years)	Setting	Description of interventions (Exercise)	Treatment duration	Outcome	Joint affected (OA)
Bennell, 2005 (Australia)	68	68 mean	Community based patients; Diagnosis: clinical and X-ray	Individual program: Strengthening, taping and massage. n=73 Control group: Sham Ultrasound control: n= 67	12 weeks	Pain (VAS) Function (WOMAC)	Knee
Tak, 2005 (Holland)	68	68 mean	Primary care and community based patients; Diagnosis: clinical	Class-based: 8 sessions of strengthening plus home program. n= 35 Control group: Waiting list. n= 39	8 weeks	Pain (VAS) Function (GARS)	Hip
Hay, 2006 (UK)	64	68 mean	Primary care patients; Diagnosis: clinical	Class program: 3-6 sessions with physiotherapist plus exercise advice. n=93 Control group: Advice/education leaflets plus telephone call. n= 94	6 months	Pain (WOMAC) Function (WOMAC)	Knee
Mikesky, 2006 (USA)	60	69 mean	Community based patients; Diagnosis: clinical and X-ray	Home program; 45 clinic sessions: strengthening with Theraband. n=15 Control group: Range of motion (ROM) exercise. n=22	12 months, 30 months follow up	Pain (WOMAC) Function (WOMAC)	Knee
Wang, 2007 (USA)	84	66 mean	Primary care and community based patients; Diagnosis: clinical	Group program: Aquatic exercise focused on strengthening. n=20 Control: Non-exercise control conditions. n=18	12 weeks	Pain (VAS) Function (VAS)	Knee or hip

Table7.4 continued

Author, Year (Country)	Gender (%) Female	Age (Years)	Setting	Description of interventions (Exercise)	Treatment duration	Outcome	Joint affected (OA)
Fransen, 2007 (Australia)	75	70 mean	Community based patients; Diagnosis: clinical	Class based program: Hydrotherapy program n=55 Control group: Waiting list. n=41	12 weeks	Pain (WOMAC) Function (WOMAC)	**Knee** or hip
Hinman, 2007 (Australia)	68	62 mean	Primary care and community based patients; Diagnosis: clinical and X-ray	Aquatic group: Lower extremity exercise, flexion/extension and walking. n= 36 Control group: Usual daily activities and medication regimen. n=35	6 weeks	Pain (WOMAC) Function (WOMAC)	Knee or Hip

7.3 Risk of bias within studies

In general the studies included in this review appeared to be of good methodological quality after assessing bias within the four domains (randomization, concealment allocation, blinding assessors and lost to follow-up) recommended by Cochrane [*Higgins et al 2011*]. Figures 7.2 to 7.5 present the proportions of studies that are of low, high or unclear risk of bias for each of the four domains for the four interventions evaluated in this review.

All of the 41 RCTs (4 advice and information RCTs, 2 simple analgesia RCTs, 4 topical NSAIDs RCTs and 31 exercise RCTs) had low risk of bias for the randomization domain. For the advice/information intervention, there was no clear information on 75% and 33% of the studies on concealment of treatment allocation and blinding of outcome assessors respectively with 25% of the studies having high risk of bias for loss to follow up (i.e. response rate at the end of treatment). All the topical NSAIDs studies had low risk of bias for all the domains. For simple analgesia, 50% each of the studies did not provide clear information on concealment of treatment allocation and loss to follow up respectively with the remaining domains having 100% low risk of bias. For the exercise intervention, the proportion of studies with high risk of bias were 23% and 16% for the blinding of outcome assessor and loss to follow up domains respectively whilst the proportion of studies with unclear information on concealment of treatment allocation, blinding of outcome assessor and loss to follow up domains were 42%, 13% and 23% respectively.

Sensitivity analyses excluding any study with high risk of bias in at least one domain was performed and the results are presented in section 7.4.5.

Figure7.2 Risk of bias within the advice and information studies

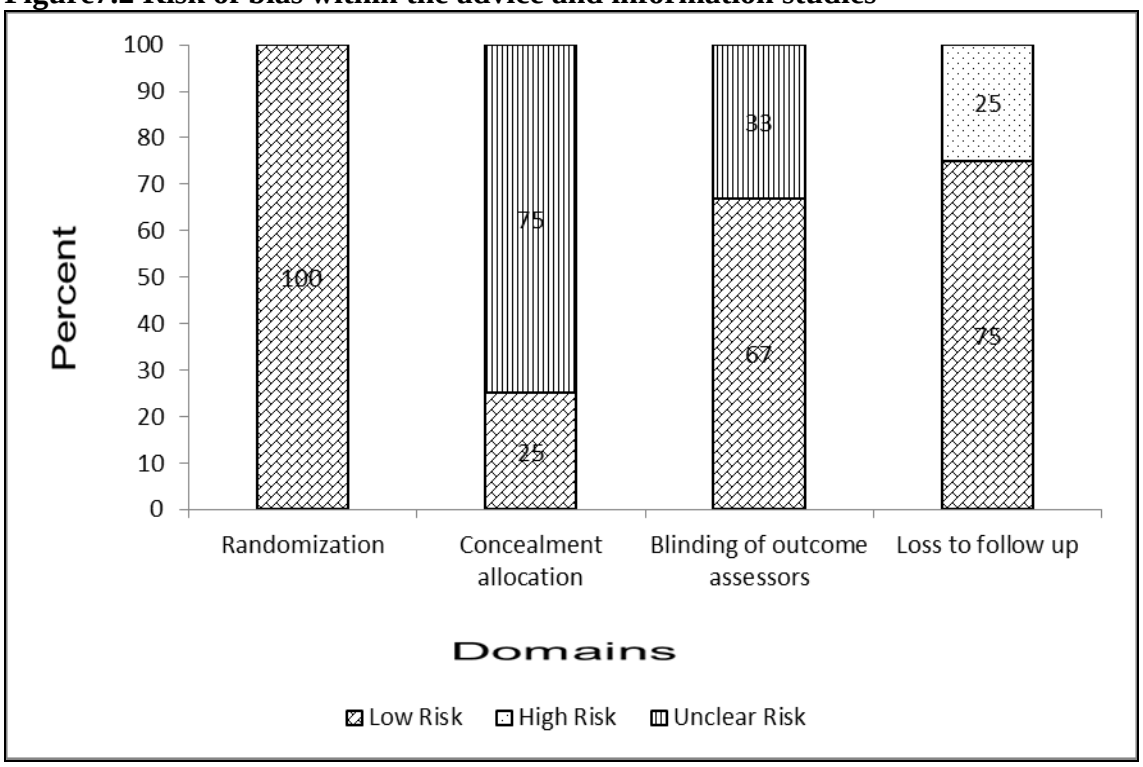


Figure7.3 Risk of bias within the simple analgesia studies

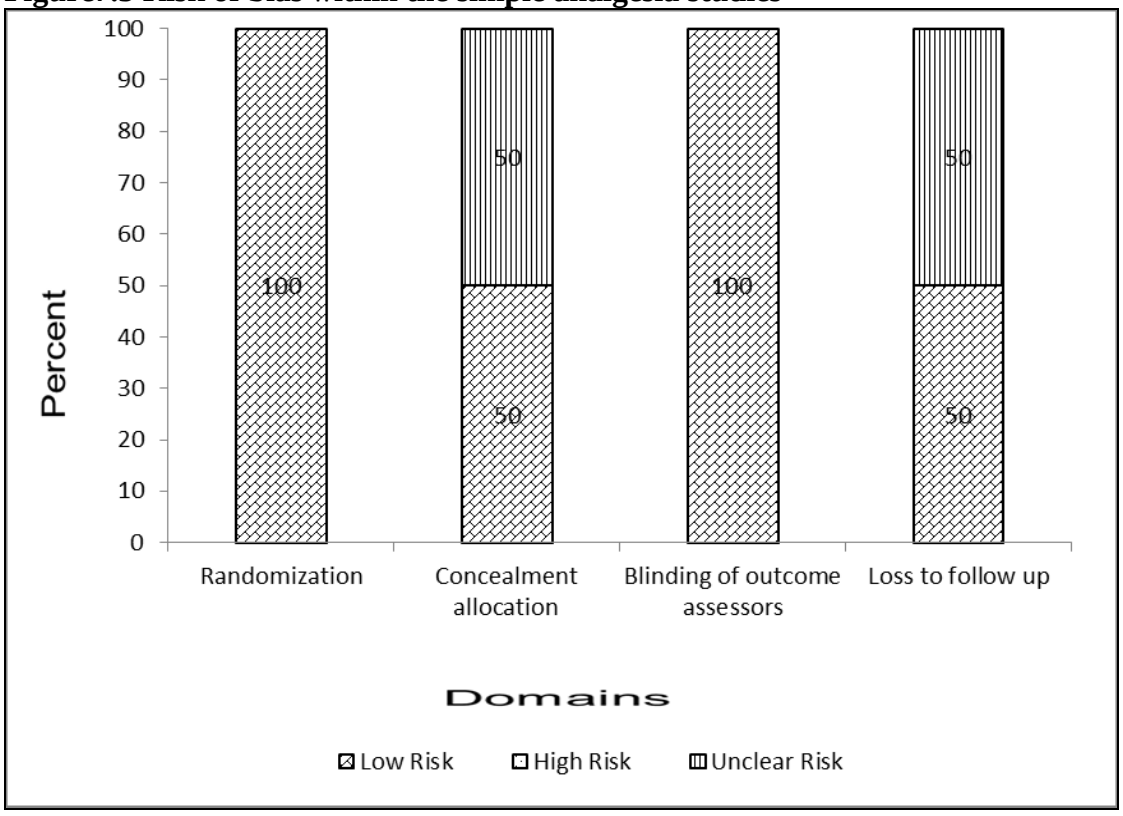


Figure7.4 Risk of bias within the topical NSAIDs studies

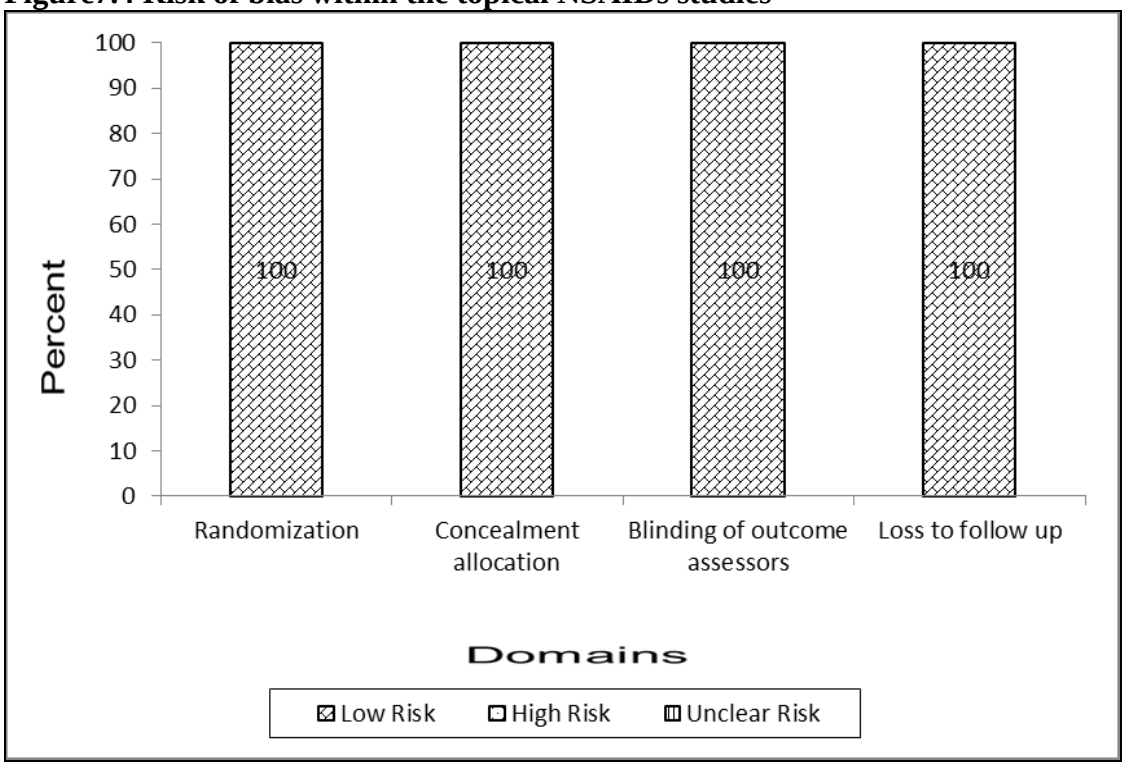
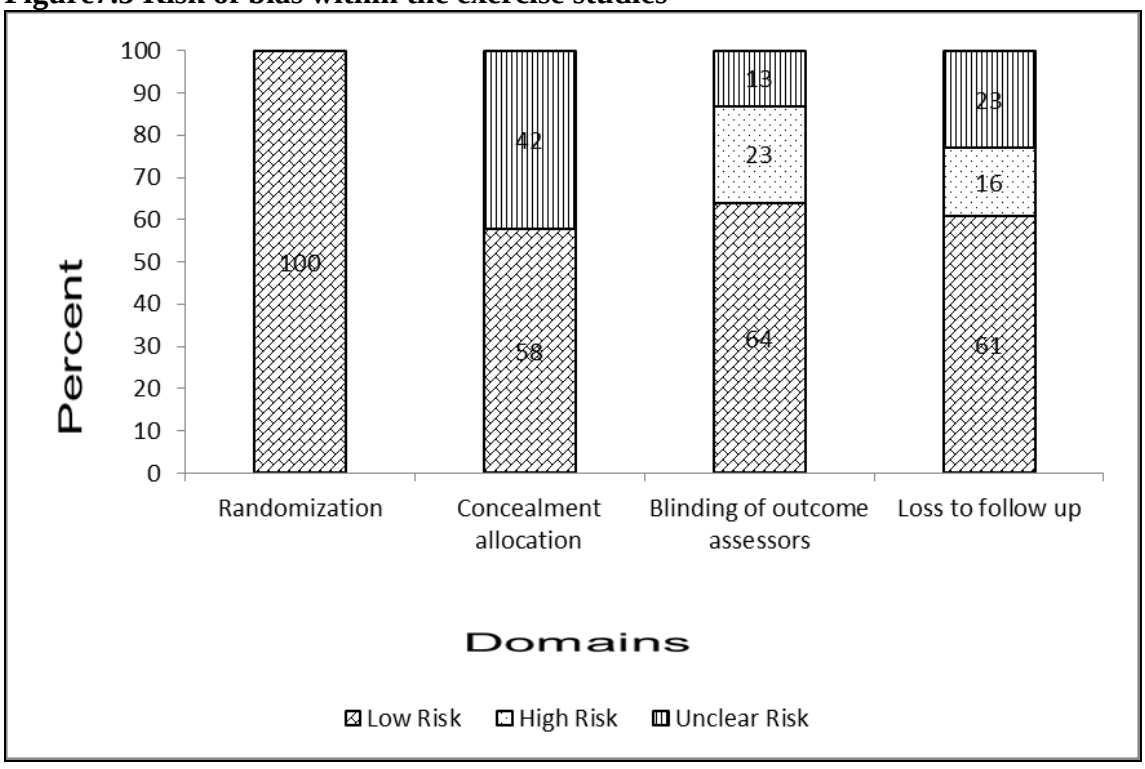


Figure7.5 Risk of bias within the exercise studies



7.4 Results

The forty–one RCTs included in this review provided data on 6,715 subjects assessed for pain and 5322 subjects assessed for functional limitation. The numbers of patients involved in pooling the effects estimates of the four interventions are stated in the respective sections below. Both end of treatment scores and change scores from baseline to end of treatment were used and where appropriate for both outcome measures (pain and function). Evaluation of the effectiveness of the combined results of the RCTs of the four interventions is as described below.

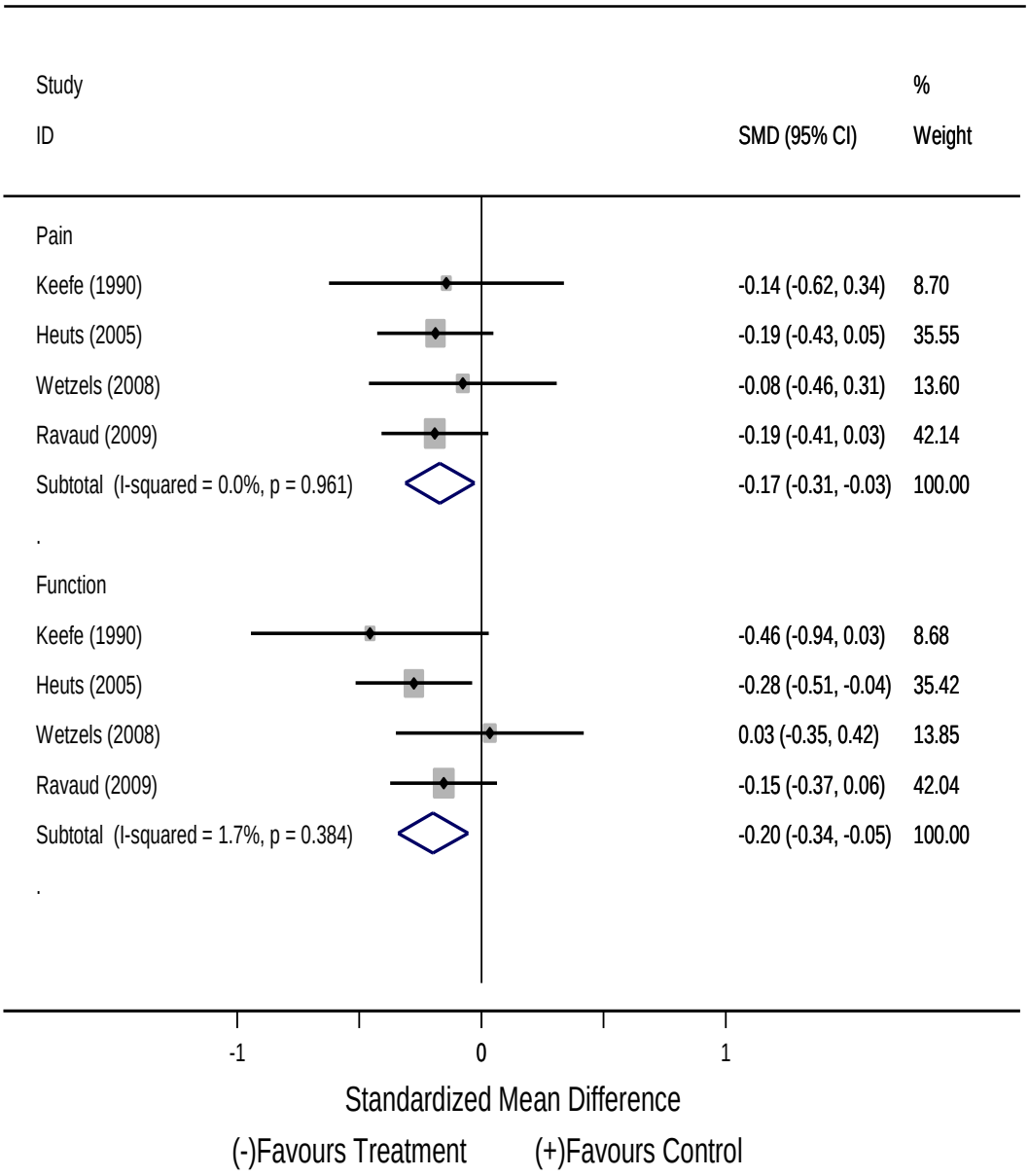
Effect estimates, i.e. the SMD and 95% CIs of each study and pooled estimates are presented in forest plots. The forest plots help to visualize the results of both individual studies as well as the pooled findings. It displays effect size estimates, represented by a square block, with their 95% confidence intervals presented by a horizontal line extending either side of the block for individual studies as well as giving the overall estimate (typically represented by a diamond). The size of the block reflects the weight (i.e. inverse variance) assigned to that study.

7.4.1 Advice and information

Figure 7.6 below shows the forest plot of four RCTs comparing advice and information intervention to suitable controls. 730 of the subjects were in the intervention arm whilst 812 subjects were in the control arm. Only one RCT [*Heuts et al 2005*] yielded a significant

finding, showing a significant improvement in function in favour of advice and information with a treatment duration of 3 months. There was no evidence of heterogeneity across the studies for either pain (I^2 = 0.0%, p-value = 0.961) or functional limitation (I^2 = 1.7%, p-value = 0.384) outcomes. When the effect estimates (SMD) of the 4 RCTs were pooled, there was an overall small, but significant, reduction in pain (SMD = -0.17; 95% CI: -0.31 to -0.03, n = 771) as well as a small significant improvement in physical function (SMD = -0.20 95% CI: -0.34 to -0.06, n= 771) in favour of advice and information compared to the control group with a mean duration of treatment of 19 weeks. The number of studies was too small to reliably assess the risk of publication bias for this intervention.

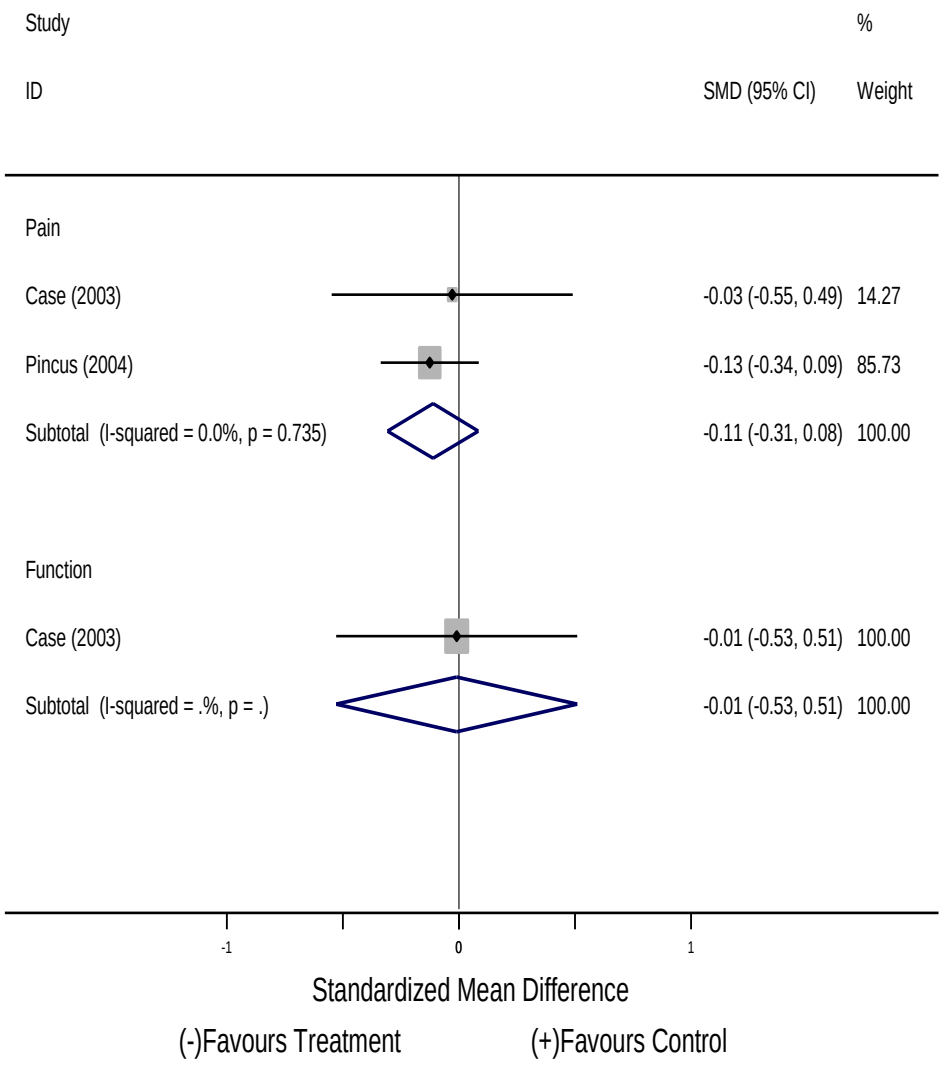
Figure7.6 Effect estimates (SMD) of advice and information for pain and functional limitation outcomes



7.4.2 Simple analgesia

Only two RCTs of simple analgesia met the inclusion criteria in this review and comprised of 229 subjects in the intervention arm and 228 subjects in the control arm. Both RCTs failed to demonstrate a significant difference in reduction in pain or functional limitation between intervention and control groups. For both pain and functional limitation, the estimated pooled effect size of simple analgesics was small and statistically insignificant: SMD = -0.11(95% CI -0.31 to 0.08, n = 400) and -0.01 (95% CI: -0.53 to 0.51, n = 57) respectively (Figure7.7) with a mean duration of treatment of 9 weeks. Since the number of studies was very small, the results have to be interpreted with caution. Also, neither evidence of heterogeneity across studies nor risk of publication bias could be assessed reliably for this intervention**.**

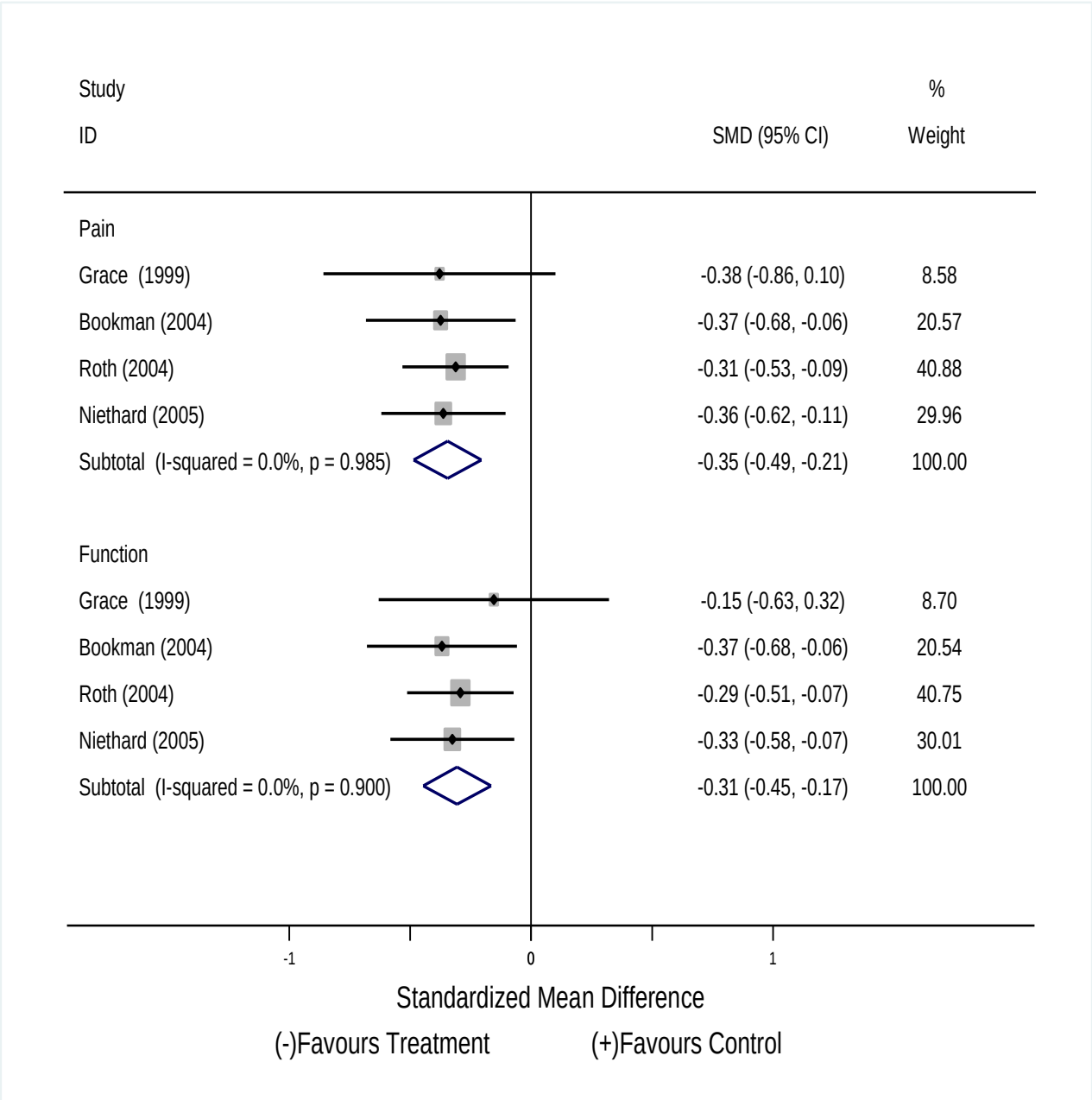
Figure 7.7 Effect estimates (SMD) of simple analgesic for pain and functional limitation outcomes



7.4.3 Topical NSAIDs

Four RCTs investigated the effects of topical diclofenac (n=795) compared to placebo (n=784), on both pain and functional limitation. The SMD of three of the RCTs was statistically significant in favour of diclofenac compared to placebo for both outcome measures, whilst the Grace et al [1999] RCT showed no significant effect of diclofenac on reduction of pain and functional limitation compared to controls. There was no evidence of heterogeneity of effects amongst the studies for either pain (I^2 = 0.0%, p=0.985) or functional limitation (I^2 = 0.0%, p=0.900). However, this has to be interpreted with caution as the number of studies considered was small (i.e. n=4). The pooled SMD estimates showed a statistically significant moderate reduction in pain (-0.35; 95% CI -0.49 to -0.21, n = 790) as well as moderate improvement in function (-0.31; 95% CI -0.45 to -0.17, n=789) among the intervention groups as compared to the controls (Figure 7.8) with a mean duration of treatment of 5.5 weeks. The number of studies was too small to reliably assess the risk of publication bias for this intervention.

Figure 7.8 Effect estimates (SMD) of topical NSAIDs for pain and functional limitation outcomes



7.4.4 Exercise

The total number of subjects in the exercise intervention arm was 4563 whilst 3896 were in the control arm. Of the thirty-one RCTs for exercise which were included in this analysis, only three showed a statistically significant reduction in pain [*van Baar et al 1998b; Deyle et al 2000; Fransen et al 2007*] (Figure7.9a) and similarly for improvement in functional limitation [*Bautch et al 1997; Deyle et al 2000; Fransen et al 2007*] (Figure7.9b). There was no evidence of heterogeneity between the studies for pain (I^2=0.0%, p=0.990) or for physical function (I^2=0.0%, p=0.726). Pooling of the findings demonstrated a statistically significant moderate reduction in pain (SMD = -0.32; 95% CI: -0.43 to -0.21, n=1389) and moderate improvement in function (SMD = -0.27; 95% CI: -0.39 to -0.16, n = 1240) among the intervention group compared to controls with a mean duration of treatment of 19 weeks. The funnel plots for both pain and physical function outcomes appear to be symmetrically shaped indicating the absence of publication bias even though one study's estimate for the physical function outcome appears to be an outlier (Figures 7.10a and b). Egger's bias estimates confirm this lack of publication bias for both pain (bias = -0.49; 95% CI: -1.41 to 0.44, p= 0.292) and physical function (bias = -1.14; 95% CI: -2.47 to 0.18, p= 0.087).

Figure 7.9a Effect estimates (SMD) of exercise for pain outcome

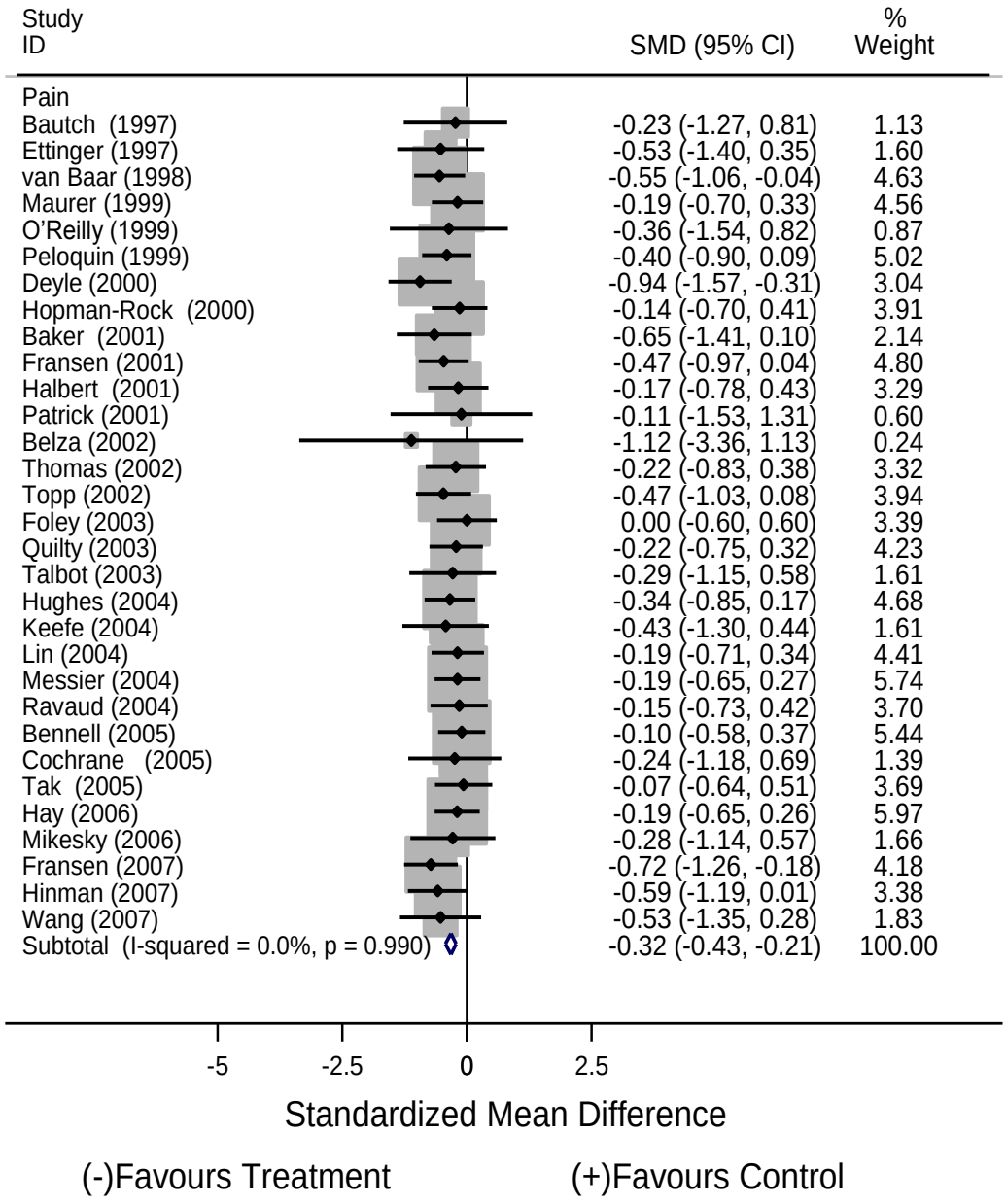


Figure 7.9b Effect estimates (SMD) of exercise for functional limitation outcomes

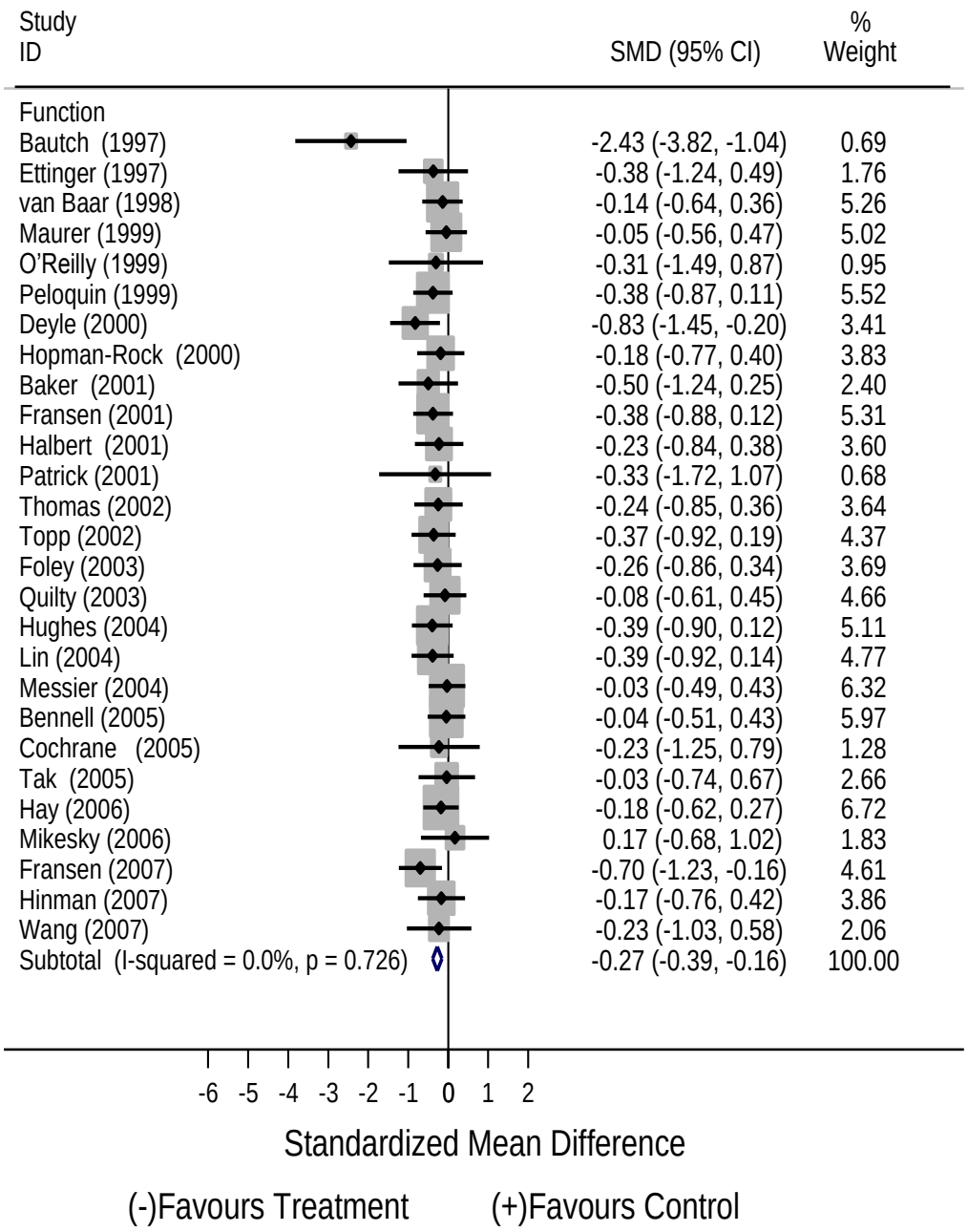


Figure 7.10a Funnel plot with 95% CI for exercise intervention for pain outcome

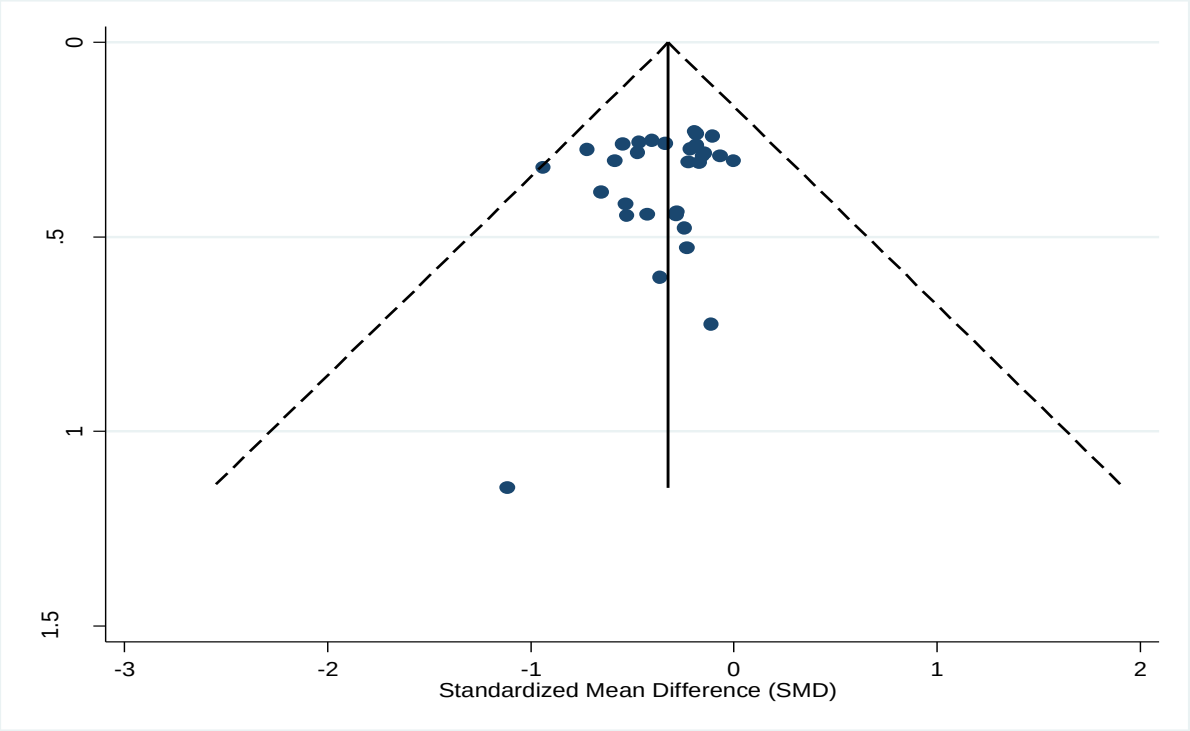
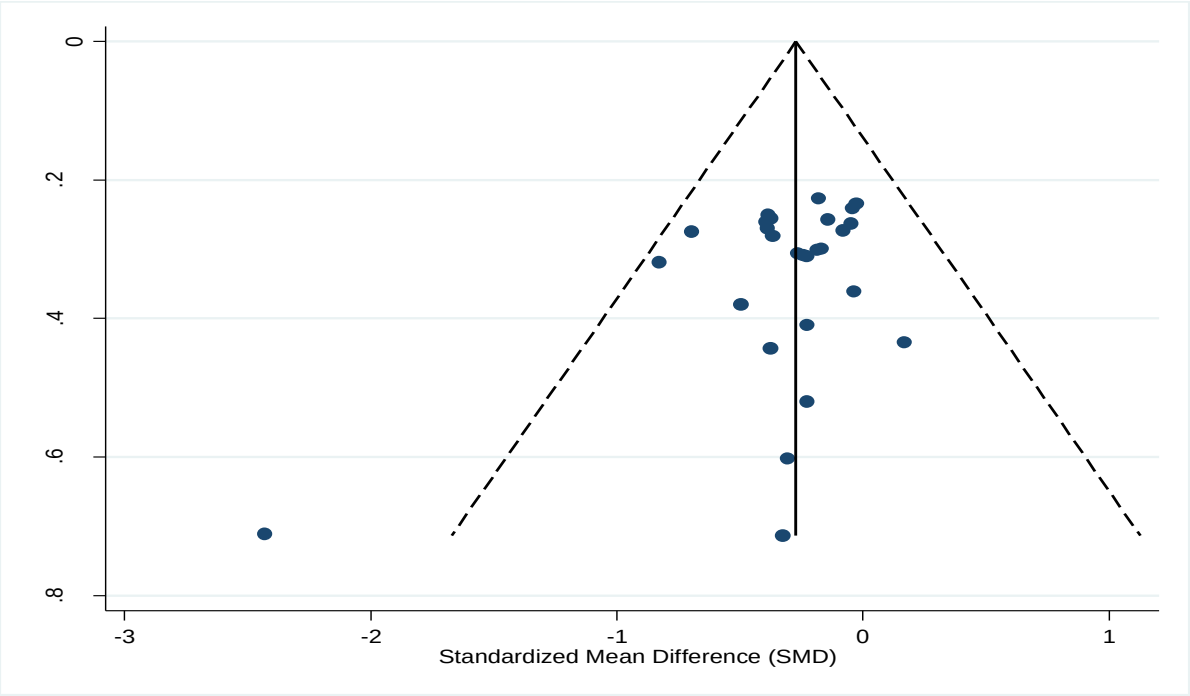


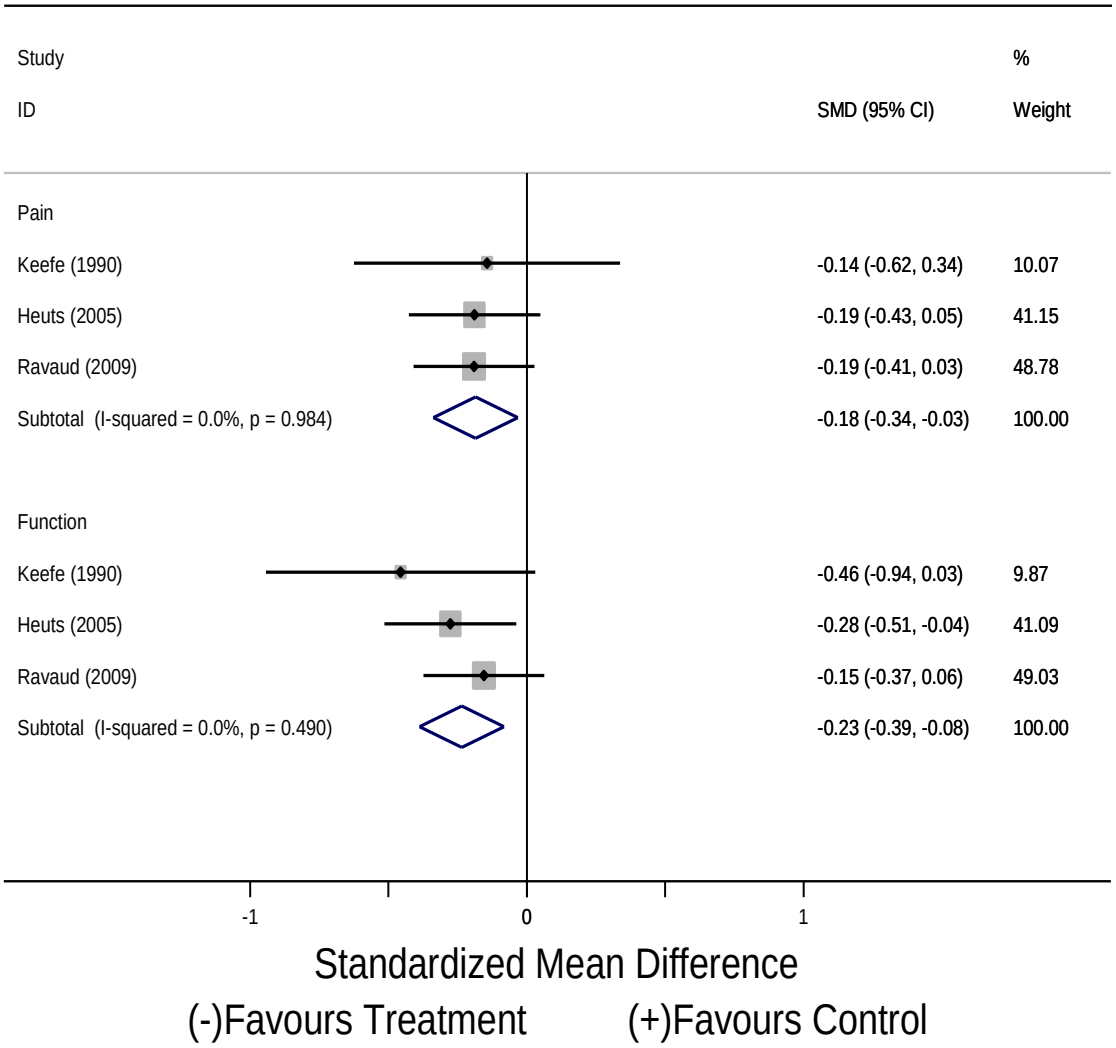
Figure 7.10b Funnel plot with 95CI for exercise intervention for functional limitation outcome



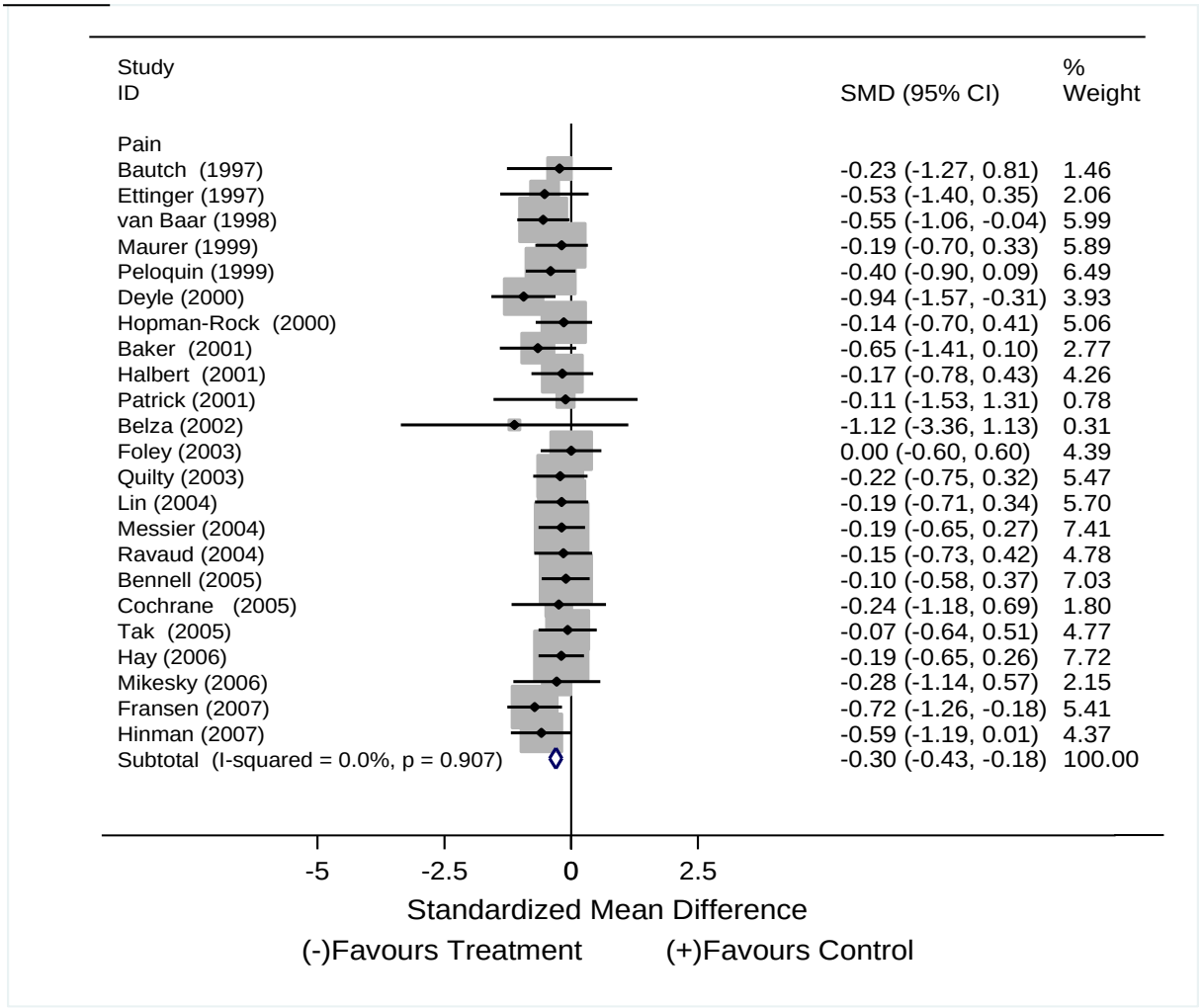
7.4.5 Sensitivity analyses excluding high risk of bias studies

After excluding studies with high risk of bias (in at least one of the domains considered in this study), sensitivity analyses were performed for advice/information and exercise intervention studies and the results showed minimal differences in the pooled effect estimates for the two interventions. For advice/information intervention, the pooled effect estimates increased by 0.01 and 0.03 for pain and functional limitation outcomes respectively whilst for exercise the pooled effect estimates decreased by -0.02 each for both pain and functional limitation outcomes. The number of studies used for the sensitivity analyses was 3 (instead of 4) for the advice/ information intervention for both pain and functional disability whilst that for exercise intervention was 23(74%) and 21(68%) studies (instead of 31) for pain and functional disability respectively. Figures 7.11a to 7.11c shows the effect estimates (SMD) and their 95% CIs for advice/information and exercise for pain and functional limitation outcomes respectively.

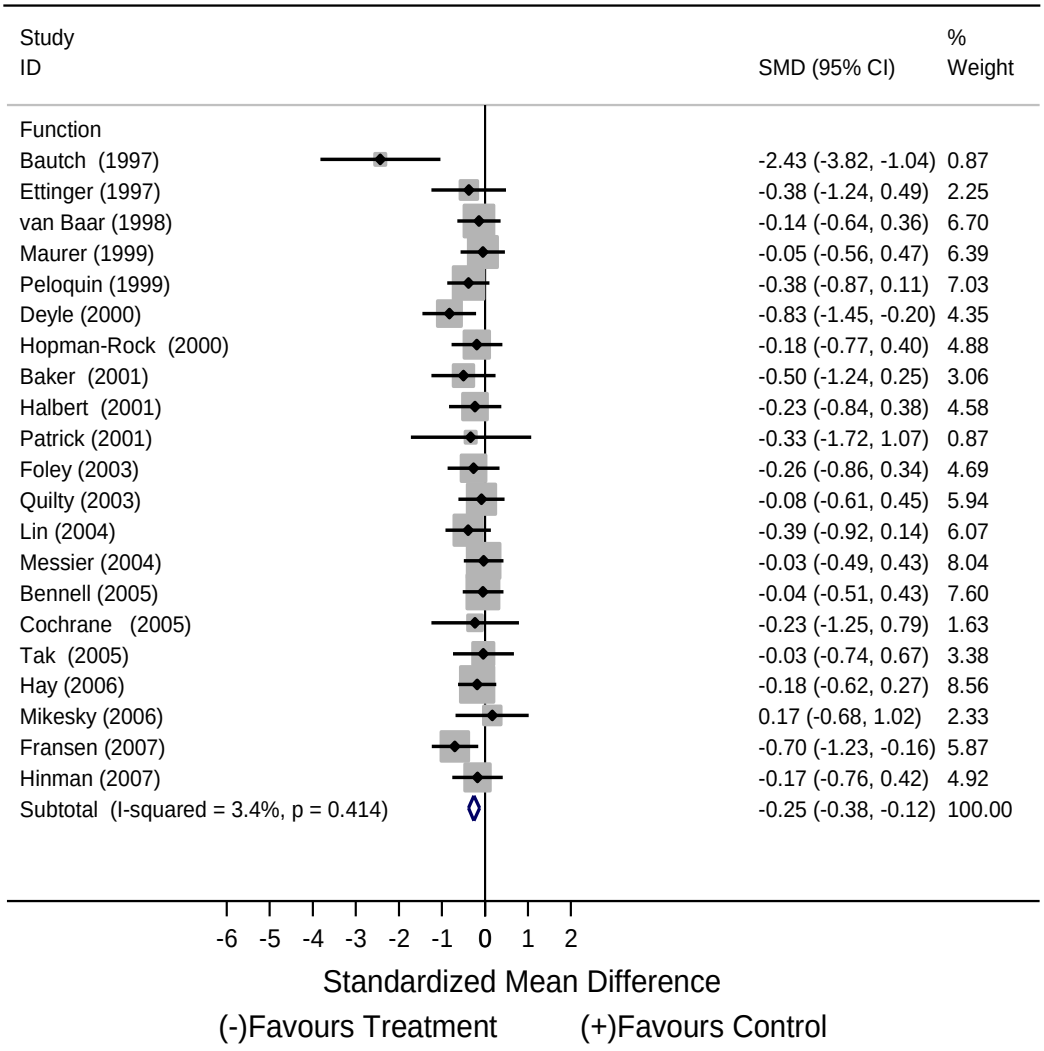
Figures 7.11a Effect estimates (SMD) of advice and information for pain and functional limitation outcomes: sensitivity analysis excluding RCTs with a high risk of bias at any one domain



Figures 7.11b Effect estimates (SMD) of exercise for pain outcome: sensitivity analysis excluding RCTs with a high risk of bias at any one domain



Figures 7.11c Effect estimates (SMD) of exercise for functional limitation outcome: sensitivity analysis excluding RCTs with a high risk of bias at any one domain



7.5 Discussion

This section summarizes the findings of the evidence synthesis and meta-analysis and compares findings with those of other studies, describing strengths and weaknesses and providing general conclusions.

7.5.1 Summary of findings

This evidence synthesis and meta-analysis involved summarizing RCTs that evaluated advice and information, simple analgesia, topical NSAIDs and exercise interventions for patients with OA. Knee and hip were the most commonly investigated joints. The topical NSAIDs studies were assessed as having low risk of bias for all domains. For the other interventions, high risk of bias was found in one advice and information study and in 8 exercise intervention studies. Lack of information was found in at least one of the domains in about half of the studies of the 3 interventions except topical NSAIDs. The findings of sensitivity analyses excluding studies with high risk of bias showed minimal changes in the pooled effect estimates for advice & education or exercise interventions.

Overall, the meta-analyses demonstrated that information/advice, topical NSAIDs and exercise interventions may give significantly more reduction in both pain and functional limitation as compared to the controls, however this finding did not hold for simple analgesia. The effect sizes obtained for the interventions were small to moderate based on Cohen's cut-off points [*Cohen 1977*].

Only the analysis of the effects of exercise was based on a sufficiently large number of RCTs (n=31). The number of RCTs on advice, analgesics and topical NSAIDs was small, and may have led to biased and imprecise effect estimates. This may not have been a problem if the number of participants in a study had been sufficiently large. The large number of RCTs included in the meta-analysis for exercise resulted in a narrower CI around the point estimate whereas the point estimate for simple analgesia had a broad CI and hence not statistically significant due to the small number of RCTs and small number of participants examined. Moreover, it was not meaningful to explore the potential influence of publication bias for the three interventions with small number of studies.

7.5.2 Comparison with other studies

The RCTs used in this review were obtained mainly from existing reviews (including those by Fransen et al 2002 and 2009, and Towheed 2006) as well as including an update of new RCTs from the Medline database. These RCTs investigated at least one of four primary care interventions of interest to this thesis. The difference between this review and the other reviews is that it focuses only on patients from primary/community care with OA at any joint (but mainly hip and knee) site as this is the proposed target population of the decision modelling study presented in Chapters 8 and 9. Existing published reviews involved subjects with only one joint affected whilst few others combined subjects with OA at any lower extremity joint such as knee and hip, in either primary care or secondary care or both. The search identified only few RCTs involving OA of the hand and foot, and these studies failed to meet the eligibility criteria.

The quantitative findings of this review are in agreement with findings of previous reviews on advice, medication or exercise in people with hip or knee OA. The magnitude of pooled effect size estimate pertaining to trials on advice and information in this study (SMD -0.17; 95% CI -0.31, -0.03) is similar to that of Superio-Cabuslay et al [1996] (SMD -0.15; 95% CI: -0.73, 0.43) for pain. However, improvements in functional disability was greater in this study (SMD -0.20; 95% CI: -0.34, -0.05) compared to Superio-Cabuslay et al [1996] who did not report significantly larger improvements (SMD -0.02; 95% CI: -0.47, 0.51) for advice and information compared to controls in their analysis of 10 RCTs involving patients from any health setting. However, the 95% CIs of the effect sizes of both studies overlap which indicates that this study's result is not significantly different from that of Superio-Cabuslay et al [1996]. The differences in the magnitude of effect estimates for functional limitation between this study and that of Superio-Cabuslay et al [1996] may be due to the differences in the settings as most primary care patients are likely to suffer early stage symptoms of OA compared to secondary care patients.

Regarding the simple analgesia (paracetamol) studies, this review showed no significant improvement in pain or physical function. However, Zhang et al [2004] found that paracetamol was more effective (SMD -0.21; 95% CI: -0.41, -0.02) than placebo in relieving pain due to OA in their meta-analysis involving 2 RCTs made up of 110 patients with radiographic OA of the knee and/or hip mainly from secondary care settings with relatively short treatment duration (mean 3.5 weeks). The reason for the difference in results between this study and that of Zhang et al [2004] may be attributed to the differences in study settings and length of follow up. Towheed et al [2006] whose study settings and patients

characteristics were similar to that of Zhang et al [2004] also found in their meta-analyses that paracetamol was significantly better ($p < 0.05$) than placebo in terms of pain reduction (SMD -0.11; 95%: -0.22, -0.01) but not for physical disability (SMD -0.04; 95%: -0.18, 0.10). Moreover, although Eccles et al [1998] compared paracetamol to NSAID (ibuprofen) in their meta-analysis and found NSAID superior to paracetamol they concluded that the initial treatment for OA pain should be paracetamol followed by an NSAID which conforms to current guidelines for managing OA [*NICE 2008*].

The non-significant results found for paracetamol in this review may be because only two relevant eligible RCTs were identified. However, the small effect estimate (SMD -0.11) found in this review was similar to that found by the above mentioned reviews which suggests that the effect of paracetamol on patients with OA may be consistent across the various settings they are recruited from.

With regards to topical NSAIDs, the results of this review agree with those of Towheed [2006] who combined 3 RCTs with mean trial duration of 8.5 weeks to demonstrate significant benefit of topical NSAIDs compared to placebo in terms of pain (SMD -0.33; 95% CI -0.48, -0.18) and functional disability (SMD -0.35; 95%: -0.50, -0.20) improvements. Also, Biswal et al [2006] found in their meta-analysis including four RCTs involving 322 patients treated for 4 weeks or more that topical NSAIDs were superior to placebo in reducing pain (SMD -0.28; 95% CI: -0.42, -0.14). Moreover, the effect estimate for pain and functional disability (SMD -0.35 and -0.31 respectively) found in this review was similar to those of the above two studies, despite their studies involving both primary care and secondary care patients.

Considering exercise interventions, this review's results reflect those of Fransen and McConnell [2008] who evaluated the effect of land based therapeutic exercise for people with knee OA on self-reported pain (SMD -0.40; 95% CI: -50, -0.30) and physical function (SMD -0.37; 95% CI: -0.49, -0.25) and found significant benefit in both outcomes after evaluating 32 RCTs comprising 3800 participants. Similarly, significant benefits were observed for both pain (SMD -0.46; 95% CI: -0.57, -0.35) and physical function (SMD -0.33, 95% CI: -0.43, -0.23) by Fransen et al [2002] in their meta-analysis of therapeutic exercise for people with OA of the hip or knee involving 14 RCTs comprising of 1633 participants. In contrast, Fransen et al [2009] showed a non-significant reduction in pain (SMD -0.33, 95% CI: -0.84, 0.17) for exercise in their meta-analysis involving 5 RCTs made up of 204 participants with OA of the hip. The effect estimate of exercise in this study was similar to these in terms of magnitude but the precision of estimates differed for both pain and functional limitation. In general, the range of exercise routines undertaken in the studies considered in this review were similar to those of the other studies mentioned above where the routines undertaken mainly includes land based quadriceps muscle strengthening involving leg raising and aerobic walking routines delivered to a small group of people either individually at home or in a group.

Recent reviews on the effectiveness of exercise compared to non-exercise control conducted by Jansen et al [2011] and Fransen et al [2010] have all shown significant decrease in pain (SMD -0.34, 95% CI: - 0.49, -0.19 and SMD −0.38, 95% CI: −0.67, −0.09) among adult patients with knee and hip OA respectively. The review by Jansen et al [2011] also showed a significant improvement in physical function whilst that of Fransen et al [2010] failed to

show improvement in function. Even though the above reviews were mostly made up of patients from secondary care settings, their findings were similar to that obtained in this review.

7.5.3 Strengths and limitations of the review

The results of this review provide current evidence (1990 to 2009) on the effectiveness of primary care interventions (advice, medications and exercise) used as first line treatment options for OA among patients generally from primary care settings. It was not possible to identify the setting of some of the studies, particularly those of simple analgesia and topical NSAIDs carried out in the USA where the organisation of primary care is different compared to the UK in that patients can have direct access to both primary and secondary care [*Bodenheimer. 2003*]. This review shows that few RCTs have been carried out in primary care, especially those on advice and medication and as such only a small number of studies could be compared to existing published reviews which indicate that majority of the evidence of these commonly used treatments in primary care is based on secondary care populations.

The main aim of this review was to carry out a more efficient evidence synthesis of the effectiveness of primary care interventions for OA by searching for existing reviews of RCTs of the four interventions for OA, and updating them with recent individual RCTs (rather than conducting a full systematic search for individual RCTs in all relevant bibliographic databases). The pooled estimates obtained from the meta-analysis were subsequently used in the construction of the economic evaluation model described in chapters 8 and 9. Since

recent reviews will include the most currently available RCTs, the decision to re-examine them and update them with more recently published individual RCTs saved time while missing few available relevant RCTs. This was an efficient approach to obtaining a reliable and valid estimate of the overall effect sizes of the interventions considered in this review.

The Cochrane Risk of Bias tool used to assess the risk of bias is simple, reliable, valid and applicable for patients with any type of disease. The risks of bias assessed in this review (i.e. randomization, concealment allocation and blinding) have been shown to potentially influence the estimates of an intervention's effectiveness in RCTs [*Colditz et al 1989*]. For example, it has been demonstrated that both RCTs and non RCTs that do not employ a double-blind design are more likely to show that a new treatment is significantly more effective than a standard treatment. Therefore sensitivity analyses were performed to assess the potential risk of bias in this review.

The quality of the studies in this review was good with respect to the risk of bias associated with randomization, which was low in all the studies. However, only about half of the studies were at low risk of bias in terms of concealment of treatment allocation, blinding of outcome assessment and loss to follow up. Also, information was lacking (unclear) on concealment of allocation and loss to follow-up in a few of the studies, making it impossible to correctly assess the risk of bias for these domains particularly for advice and information and simple analgesia studies. In this case relating risk of bias to the actual design and conduct of studies can be difficult since reports may not cover every detail of the way a study is carried out because of poor reporting or journal restrictions regarding word counts.

Only one reviewer assessed the risk of bias for the studies included in this review even though at least two reviewers are recommended [*Liberati et al 2009*] to perform such assessment in order to reliably determine or judge the level of bias in each study. It appeared not feasible to perform a second assessment of risk of bias within the timeframe of this study, and the reliability of the judgment observed for some of the studies may be questioned. Although this study did not aim to focus exclusively on high quality studies but rather on all available relevant studies, sensitivity analyses were performed for advice/information and exercise omitting studies with high risk of bias and the results showed very little differences in the findings.

Data collected on various characteristics of the individual studies revealed that they differed in terms of joints affected, duration of treatment, gender, age and number of patients per treatment arm, yet, the findings showed that heterogeneity was not apparent among the studies for any of the four interventions considered in this review. Visual inspection showed the effect estimates of the individual studies to be very similar and their CI overlapping. The number of studies used to pool treatment effect was small for all interventions except exercise and the method used to assess heterogeneity (i.e. I^2) is dependent on sample size of the studies in a review [*Rucker et al 2008*] - hence, care must be taken when interpreting the results. An alternative way of assessing heterogeneity is the use of tau^2 measure which refers to the underlying variability between studies in a review which was proposed by Rucker et al [2008] as it is not influenced by either the number or size of studies in a review. However, the measure is not commonly used as it does not propose cut-off value to use to decide when one should pool or not to pool studies together in a review hence it was not used in this review.

Both change scores and end of treatment scores of outcome measures were combined to calculate the effect estimates of the interventions. The advantage of using change scores of outcome measures is that it adjusts for the differences between groups at baseline. Even though RCTs incorporate randomization in their study design, differences between treatment arms may occur particularly in small sample studies. Given that change scores were used by about half of the studies included in this review this may have reduced the potential influence of baseline variability on the pooled estimates of effect sizes of the interventions. This study assumed that duration of treatment is equal to duration of follow up since only end of treatment scores were used to pool effect estimates for the treatments and also because the majority of the studies did not followed up their patients further after treatment. The advantage of this assumption is that majority of the data on the outcome is available to pool the effect estimates whereas data on long term follow up studies are usually not complete due to loss to follow up.

The number of studies was not large enough to examine publication bias for three of the interventions. The analysis of exercise interventions included a large enough number (31) of studies, and the results showed absence of publication bias although there were a few small trials ($n<30$ per arm) among the exercise studies. However, non-published trials and conference proceedings were not searched for in this review. If a lot of such studies were excluded it could result in bias of the findings (i.e. overestimation of the effect estimates) due to potential selection of studies reporting positive findings only. The majority of RCTs used in this review were identified from existing reviews and most of them searched unpublished data but did not identify any, thus it is not likely that this study have missed important

unpublished studies and hence unlikely to affect the overall effect estimates. However, the evidence that studies with positive findings are more likely to be published is mixed as some studies [*Abbot and Ernst 1998, Olson et al 2002*] have shown no association between publication of submitted articles and positive study outcomes whilst others [*Mahoney 1977; Ernst and Resch 1994*] on the other hand, have shown that peer reviewers are more likely to recommend in favour of publication if the results of a study is positive.

Only studies published in the English language were considered in this review. It is envisaged that the impact of language bias is minimal, even though it is difficult to forecast in what precise instances such exclusion may bias review findings. Non-English studies are rarely indexed in the two databases searched in this review and only one such study was identified and subsequently excluded. The importance of non-English language trials are not easy to predict in reviews and as such there is mixed evidence regarding the impact of their exclusion in reviews. For example, Jünia et al [2002] found that excluding trials published in languages other than English has little effect generally on the pooled effect of a treatment. On the contrary, Egger et al [1997] showed that the use of only trials published in English language is likely to introduce bias in reviews and meta-analyses as researchers are more likely to publish their studies in English-language journals if their results are statistically significant.

The duration of the treatment period in the studies considered in this study was variable and ranged on average from about 1 month for topical NSAIDs studies to 4.5 months for both advice/information and exercise studies with that for simple analgesia being 2 months. Since the effect estimates of the four primary care interventions were used as inputs for the

decision modelling study which used 3 monthly time cycles to reflect the effectiveness period of the interventions considered, it would have been ideal if the studies combined to estimate the overall effect sizes of the respective interventions were of similar or slightly longer length of time. This would make the overall effect estimates of the interventions more appropriate for the decision model. Given that the average duration of treatment of the topical NSAIDs and simple analgesia studies were very short and that the full effect of those interventions may not be realized, hence it is likely that their effect estimates may have influenced the decision modelling study results by either under or overestimating the findings. Those studies with longer duration of treatment are also likely to lose some of their patients to follow up and this is likely to bias the findings of this study.

7.5.4 Conclusion

The results show that access to advice and information, use of topical NSAIDs and the performance of regular exercise can result in small to moderate improvement of both pain and physical function compared to control treatments in primary care patients with OA related to the hip and/or knee. On the other hand, paracetamol does not show large or significant improvements in pain and functional disability among these patients. However this may be the result of insufficient evidence available to ascertain a more accurate estimate of effectiveness of paracetamol.

The estimates of the current overall effect size estimates of the four primary care interventions for OA (advice and information, simple analgesia, topical NSAIDs and

exercise) were used to populate the economic evaluation model developed in chapters 8 and 9, which aimed to evaluate the cost effectiveness of two different approaches to delivering optimal care interventions (namely stepped care and one-stop-shop care) compared to usual care for adults with OA.

Part 2B

Chapter Eight

Modelling cost-effectiveness of optimal primary care for OA: Background and methods

The intent of this part (2B) of the thesis is to provide a background to the economic evaluation study, describing the definition, rationale and types of economic evaluation and the choice as well as the composition of the two optimal packages of interventions used in this study. Also, the definitions, rationale and concerns for using decision models are described. The methodology used to develop the decision model (Markov model) and the results of the cost-effectiveness analysis of two optimal approaches to delivering care for OA in primary care (stepped care and one-stop-shop care) compared with usual care are described in detail.

8.1 Introduction

The decision model carried out in chapters 8 and 9 was developed with the overall aim of evaluating the cost effectiveness of two ways of delivering optimal primary care interventions compared to usual care among adults with OA. The ways of delivering primary care interventions considered in this study were stepped care and one-stop-shop care, and they contained four of the core primary care interventions for OA namely advice and information, paracetamol, topical NSAIDs and exercise. These have been described previously in the evidence synthesis and meta-analysis study in chapters 6 and 7. These optimal primary care interventions were adopted in this analysis because they are recommended by OA expert groups such as NICE (*NICE 2008*) and EULAR [*Zhang et al 2005 and 2007a*] as core interventions for managing OA. In addition, they were suggested in a consensus meeting with clinicians and OA researchers at the Arthritis Research UK

Primary Care Centre at Keele University. Both expert groups also agree that OA should be managed with several interventions.

For the stepped care approach, hypothetical participants were offered the four primary care interventions in a stepped fashion (in a supposedly ascending order of strength) where all participants start with advice/information and paracetamol interventions combined and those whose condition fails to improve over a period of 3 months move onto topical NSAIDs followed finally by exercise. The stepped approach was chosen because it will help provide people with OA with about the right amount of intervention they require given that the extent of treatment offered them depends on how well they have responded to the preceding intervention.

With regards to the one-stop-shop care approach, participants were offered all the four primary care interventions mentioned above at the same time with the aim of realizing optimum treatment outcomes for participants. However, participants whose condition fails to improve are returned to usual care. As all the participants are allowed to receive optimum care irrespective of their symptoms, participants with mild symptoms may not need such an intensive package of interventions and hence may lead to unnecessary treatment and health care costs.

For usual care, the medication data used in the Hurley et al [2007] study was used in this model where participants were potentially offered treatments according to the severity of their symptoms. As such, the interventions offered covered not only the four primary care interventions mentioned above but included other stronger medications such as opioids, oral

NSAIDs, etc, but did not include surgery as the model was developed to be applied to primary care patients. This care was chosen as the active control since it is the most appropriate comparator in economic evaluation, as it is important to determine the cost-effectiveness of alternative treatment options with what patients currently receive.

Detail descriptions of the above mentioned interventions are provided in section 8.7.4 whilst their advantages and disadvantages are described in the discussion section 9.3 in the next chapter. The specific objectives are presented in the next section, followed by the detailed description of the inputs used and the relevant procedures followed to develop the decision model.

8.2 Objectives

(i) To determine the cost effectiveness of stepped care and one-stop-shop care interventions for managing OA compared with usual care.

(ii) To carry out deterministic sensitivity analyses to compare the findings with that of the base case. For this, the important assumptions which were adopted during the construction of the decision model were varied to examine if the variations led to changes in the primary findings of this study.

8.3 Definition, rationale and outcome of economic evaluation.

Economic evaluation is defined as "the comparative analysis of alternative options (courses of action or interventions) in terms of both their costs and consequences" [*Drummond et al 2005, p9*]. This implies that any economic evaluation must aim to identify, measure, value, and compare the costs and consequences of the options it is considering. Alternative options are the various ways in which health care resources can be used to improve the wellbeing of the target population – this includes pharmaceutical, surgical, screening and health promotion interventions. Healthcare costs comprise of the value of service provided by clinical and other healthcare staff, buildings and healthcare equipment and medications. Costs can also be collected from a broader perspective, e.g. non-health service resources used to provide healthcare such as time spent to care for patients by family members.

Consequences refer to changes in a person's wellbeing which may be either positive (good health) or negative (ill health). Even though a clinical health outcome such as pain or functional limitation for OA can be adopted to measure such a change, the most commonly used health outcome in economic evaluations is the quality adjusted life year (QALY) [*Weinstein and Stason 1977*]. As the QALY is a validated, comprehensive and reliable measure of quantity and quality of life, it is recommended as an appropriate outcome measure in economic evaluation of healthcare [*NICE 2004; CCOHTA 1997*] and can be used to compare cost effectiveness of several interventions in different populations and disease areas.

QoL involves all aspect of an individual's life and not just the primary clinical outcome of a particular disease such as level of pain, and many factors can either positively or negatively impact a person's life [*Torrance 1976*]. Health related QoL which includes an individual's physical function, emotional status, level of pain and social well-being, is an important aspect of QoL.

In economic evaluation, the standard and valid measure commonly used for measuring health-related QoL is utility, which is the preferred or subjective level of wellbeing that people experience in different health states [*Drummond et al 2005; Torrance 1987*]. A health state utility is a cardinal number anchored on a scale of 0 (death) to 1 (full health), with negative values representing states worse than death (e.g. persistent vegetative or being in coma) and are commonly measured indirectly with instruments (questionnaires) such as EuroQol 5 Dimensions (EQ-5D) [*Brooks 1996*] and Short Form health questionnaire 6 Dimensions (SF-6D) [*Brazier et al 2002*]. QoL (measured as utility) is multiplied by quantity of life (measured in years) to produce QALYs. Some of the techniques available for directly measuring health state utility values include visual analogue (rating) scales, standard gamble and time trade off. Economic evaluation usually estimates an incremental cost effectiveness ratio (ICER) which denotes the cost per additional QALY gain for a new treatment compared to a control treatment and is calculated as the difference in costs divided by the difference in QALYs between two treatments [*Drummond et al 2005*].

Rationale for economic evaluation

As health care resources are scarce, choices have to be made to ensure that the greatest benefit is obtained from them. It is important to carefully consider all factors (treatment options, adverse effects, cost, etc) available so that relevant options or alternatives can be evaluated appropriately and meaningfully in order to inform decision making concerning the commissioning of available resources to one treatment instead of another. For example in order to reduce the morbidity of a chronic condition such as OA, it would be appropriate to evaluate all individual options separately, as well as appropriate combinations of these, which may help identify the most efficient management option(s) to be adopted.

Another justification for economic evaluation is to maximise benefit from a limited budget. This can be achieved when health care systems which control the introduction of new interventions for a particular disease choose those interventions that are cost effective compared to other treatment options for their patients. Because of this, in England and Wales for example, new health interventions are assessed by NICE to determine if they are cost effective before they are recommended to be used for patient care. NICE recommends that for a new treatment to be cost effective its cost must be below £20, 000 per QALY [*NICE 2008a*]. For example, for the treatment of OA, less expensive drugs such as paracetamol are preferred to drugs such as glucosamine which is more expensive [*BNF 2010*] with similar efficacy but may have side effects [*Towheed et al 2005*].

8.4 Types of economic evaluation

Four main types of economic evaluation are available for use in health care: cost effectiveness analysis (CEA), cost utility analysis (CUA), cost minimization analysis (CMA) and cost benefit analysis (CBA). These techniques are similar in the sense that they examine two or more interventions by comparing their costs and effects (benefits). The key difference between these techniques is how their outcomes (consequences) are measured and the choice of the appropriate analysis depends on the question being asked. In this study a CUA was adopted in a Markov model to evaluate the cost effectiveness of stepped care and one-stop-shop care compared with usual care. Below are brief descriptions of the four types of economic evaluation.

8.4.1 Cost effectiveness analysis

CEA is usually applied to examine the differences in cost and outcomes (effects) of health care interventions in which the outcome of interest is measured in its natural units e.g. pain, physical function, blood pressure, cholesterol, etc [*Robinson 1993c*]. It can be used to compare estimates of cost-effectiveness in different conditions such as OA and Rheumatoid Arthritis (RA) when a similar common natural unit of measuring outcome measure (such as pain score or physical function score) is used for both conditions. However, the technique cannot be used to compare conditions such as OA and diabetes where the types of outcome differ – for example pain and blood glucose levels.

8.4.2 Cost utility analysis

CUA is currently the most widely used economic evaluation method for making decisions about allocating health resources [*Drummond et al 2005*] hence it was used as the method of analysis in this study. It was also used because the disease under investigation in this study (OA) has a direct impact on QoL. CUA is used to compare different interventions in terms of their costs and outcomes where the outcome measured is based on utility values and survival with the aim to overcome the single component limitations of a CEA. The advantages associated with this method are that the estimation of cost per QALY gained enables the comparisons of different treatments across different disease areas. Also, life-enhancing treatments (such as therapies) can be compared with life-saving treatments (such as surgery) for different conditions [*Drummond et al 2005*]. However, the main disadvantage of this technique is that the derivation of utility is subjective as it combines patients' preference scores of various aspects of life (e.g. physical, emotional and social well-being) and may not always reflect a person's true health state [*Drummond et al 2005*]. This method was used in this study as it would help to compare OA with other conditions which will enable health services decide which interventions to invest in.

8.4.3 Cost minimization analysis

This technique is generally used when the interventions under consideration are assumed or expected to be similar in terms of their effectiveness [*Drummond et al 2005*]. As a result, the cost of each intervention is evaluated and the cheapest is adopted. For instance, Seferlis et al

[2000] used CMA to compare the costs of GP programme (control), manual therapy and intensive training programme for the treatment of acute low-back - all the procedures were found to be similar in terms of their effectiveness but the GP programme was found to be the cheapest programme compared to the manual therapy and intensive training programmes. However, CMAs are rarely used in practice given that very few programmes/procedures tend to be equally effective [*Drummond et al 2005*] but have been suggested to be used in equivalence trials [*Briggs and O'Brien 2001*].

8.4.4 Cost benefit analysis

Cost benefit analysis is where both the costs and effectiveness of different interventions are measured in monetary units [*Robinson 1993e*] and the decision regarding the intervention to adopt is based on assessing whether the value of the benefits exceeds the costs. The method allows health care costs and benefits to be compared not only with other health-related costs and consequences but also non-health-related costs and consequences and thus helps to ascertain if a treatment delivers an overall gain to society. It could be used to examine a single intervention even though the alternative usual care or receiving no intervention is usually implied [*Drummond et al 2005*]. The drawback of this technique is that it is not easy to estimate the monetary value for health care consequences such as pain, physical function or quality of life.

8.5 Methods for undertaking economic evaluation

There are two main methods for conducting economic evaluations in health care: trial-based analysis and model-based analysis. The trial-based method is where economic evaluations are conducted alongside RCTs where individual patient data are used by researchers to estimate cost effectiveness [*Briggs et al 2006*]. This is done by collecting economic data (such as cost and quality of life) alongside trial data (i.e. treatment effects and patients characteristics) to perform cost effectiveness analysis - the process sometimes called "piggybacked" economic evaluation. The advantages are that collecting both cost and outcome data from the individual patients is simple, cheaper and time saving.

However, trial-based economic evaluations have several limitations. For instance, a single RCT may not be able to include and compare all available options, incorporate all the relevant inputs (costs and consequences) or be carried out over reasonably long period of time to examine differences in costs and consequences [*Sculpher et al 2006*]. As a result, model based methods (decision analytical models) are used as an alternative economic evaluations method to overcome the limitations of trial-based analysis. The background and rationale for using decision analytic modelling in this thesis are described in the next two sections.

8.6 Definition, rationale and concerns of decision modelling in economic evaluation

Decision models in economic evaluation are any mathematical structure which represent the evaluations of options in terms of health and economic outcomes (cost and consequences) of patients [*Drummond and McGuire 2001*]. The technique originated from operational research and was first used by the medical research community in the 1970s [*Corner and Kirkwood 1991*]. It was defined by The International Society for Pharmaco-economics and Outcomes Research (ISPOR) Task Force on good practice in modelling studies as "an analytic methodology that accounts for events over time and across populations, that is based on data drawn from primary and/or secondary sources, and whose purpose is to estimate the effects of an intervention on valued health consequences and costs" [*Weinstein et al 2003, p4*].

Decision models are considered powerful tools for decision making in economic evaluation as they are capable of providing combined evidence on clinical and economic outcomes from a wide range of sources to help inform decisions about clinical practices and healthcare resource allocation. Such information is very important for guideline committees and decision making bodies such as NICE. As a result, NICE often uses decision models in their economic evaluations to support their guidelines in terms of recommending the use of new treatments or programme options for a particular patient group.

Rationale for decision models

Decision models are most suitable for extrapolating results from trials over a long period of time and also when there is some amount of uncertainty about the appropriate healthcare strategy to undertake [*Hunink et al 2001*]. They can be used for a condition such as OA by providing an appropriate structure which will reflect the possible outcomes that patients with OA may experience over longer periods of time, and show how the interventions being evaluated may impact on these outcomes. The circumstances under which decision modelling may be more appropriate and preferred compared to trial-based economic evaluation analysis are as described below.

The first advantage of decision modelling is that it can allow for extrapolation of findings to be made over a suitably long period of time to reflect all the important differences between health care programmes in terms of cost and consequence. Often the appropriate time horizon is the patients' lifetime particularly for interventions with a potential effect on mortality. Single source patient level data studies such as RCTs rarely follow patients long enough to assess outcomes using a lifetime horizon. For example, the duration of assessments in most trial-based cost effectiveness studies are often short, but can be longer than about 12 months [*Whitehurst et al 2011; Hurley et al 2007*].

Secondly, decision modelling facilitates the comparison of all possible relevant options. By comparing all the relevant options that are applicable in practice, it provides decision makers with the best information regarding the most appropriate (cost effective) option to adopt. Although RCTs are methodologically robust to gain information on the efficacy of an

intervention (i.e. use procedures such as randomization and concealment allocation in their design which helps to minimize bias in their findings), they are often not capable of comparing all alternative options relevant to economic appraisal. For example, the options considered in this study were a combination of several primary care interventions for OA applied at varied stages of OA progression which are applicable in clinical practice which may not be easily incorporated in RCTs for practical reasons including cost, loss to follow up and complexity of delivering interventions.

Thirdly, decision models are able to synthesize all relevant evidence from several sources (studies) to provide guidance for decision makers. For example, meta-analysis can be used to estimate the overall effectiveness of the options under consideration for a particular condition such as OA to populate a decision model. This technique was used in this study and is described in chapters 6 and 7 in this thesis.

Fourthly, the structure of decision models helps to appropriately link intermediate endpoints (e.g. cholesterol level) to final endpoints (e.g. stroke or death) to aid decision making particularly for chronic health conditions in the long term. Given that RCTs generally have short time frames, they are usually restricted to evaluating short term or intermediate clinical outcomes which may not include long term clinical outcomes such as mortality.

Finally, decision models have the capacity to incorporate all possible health states (such as absence of pain, mild, moderate and severe pain) which reflect the natural history of a health condition such as OA as well as apply varied/complex interventions at varied time periods when estimating the cost and effectiveness of optimal management options.

Concerns about decision models

Even though decision models can be used to model complicated scenarios to aid decision making, they have limitations regarding the way they are employed. For instance, they combine data from different sources which may be subject to various forms of bias in terms of patient selection and method of analysis employed in these different sources. They are sometimes condemned as lacking transparency since it is often not obvious how some of the parameters used from the different data sources are obtained or calculated. Also, they involve the incorporation of assumptions (e.g. classifying no and mild pain people as the same), which may be unrealistic about clinical practice and/or the natural history of the condition under consideration [*Briggs et al 2006*].

8.6.1 Types of decision models

The type of model used depends on the nature, time frame and complexity of a decision problem. The most common decision models employed in healthcare are decision trees and Markov models although there are alternative modelling approaches such as individual sampling simulation, discrete event simulation and dynamic models. Below are brief descriptions of these various types of approaches.

Decision tree

A decision tree is the simplest structure of a decision model. The tree typically starts at the root (decision node) which represents the decision problem, in terms of alternative treatment options to be addressed. Out of the decision node come a range of pathways (chance nodes) which describe the alternative options being considered. From the chance nodes are a series of branches which represent specific events a patient may experience. The probability of a patient passing through each branch is determined and these must sum up to 1. At the end of each branch is a terminal node which records the costs and outcomes for each of them. The overall cost and outcome for each intervention is usually estimated by the "roll-back" procedure in which each cost and outcome of the interventions are weighted by the probabilities and summed. This type of model is suitable for analysing acute diseases and one-off events which have a short and fixed time horizon. Even though the method may be used for this study, it would result in the model structure being too complex and complicated to apply as all patients' pathways must be shown including their respective probabilities, costs and utilities for the duration of the study. Hence, this approach could not be used for this study as it involved a chronic condition (OA) with a long time horizon where patients would be moving between different health states.

Markov models

Markov modelling is a stochastic (random) process where subjects move from one distinct state to another or stay in the same state based on transition probabilities. The technique is commonly used in economic evaluation of health care to model the prognosis of a clinical problem with on-going risk such as OA. The technique enables subjects to move between health states over a specified period of time divided into time cycles of equal length up to the end of a study. The time cycle used usually depends on the decision problem and the natural history of the disease. A Markov model in which the transition probabilities are constant over time is known as Markov chain and this was used in this study as it reflects the natural history of OA and is also simple to apply in a model.

The technique has an important assumption known as the 'Markovian assumption' which describes its memory-less property. This implies that a patient's previous health states are not considered and hence patients in a particular health state are treated as homogenous irrespective of the state they came from or how long they have stayed in previous state(s). This assumption can be relaxed by creating additional health states to reflect the clinical history of a condition and hence lead to a more complex and complicated model structure.

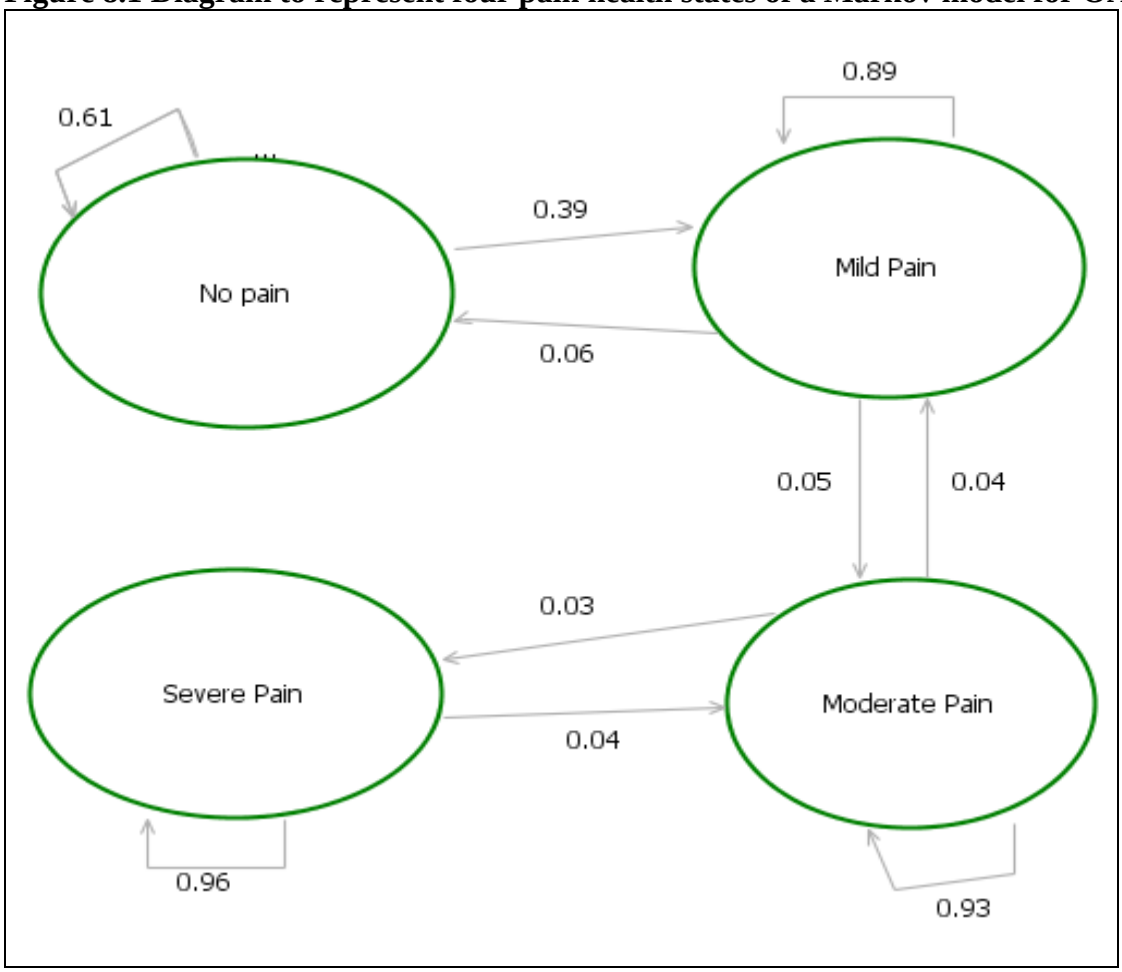
Costs and utility values are assigned to each health state. In addition, a one-off cost or reduction (or improvement) in utility can be assigned when an event occurs at a particular branch of a model. The overall costs and QALYs are obtained by multiplying (weighting) the costs and utility values by the time spent in each health state and then summed across all the health states for each of the options being considered.

Simulation is generally used for Markov modelling and it is the process of modelling random event(s) where the simulated outcomes reflect real world outcome. The most commonly used simulation method is Monte Carlo simulation which involves random sampling from probability distributions. Markov models are normally run in two main ways either by using cohort simulation or individual subject simulation (first order simulation).

For cohort simulation, the usual assumption is that the whole cohort can begin at time zero in one health state or where necessary the cohort can be distributed between all the health states in a model where their movements between health states are based on constant probabilities. For instance, consider figure 8.1 below which illustrates the transition of OA patients among four pain health states (no pain, mild, moderate and severe pain). Each oval represents a health state, the arrows represent transitions and the numbers along the arrows indicate the transition probabilities. The probabilities of the transition arrows which emerge from any state sum up to 1. For example the probability of moving from no pain health state to mild health state is 0.39, which indicates that at each cycle a constant proportion of 0.39 patients will move from no pain health state to mild health state whilst a proportion of 0.61 will remain in no pain health state using the cohort method. This method is simple, fast to run, transparent as it shows the proportion of the cohort in each health state at each cycle based on constant transition probabilities and appropriate to apply to people with a chronic condition such as OA, hence cohort simulation was used in this study.

The model was constructed using TreeAge Pro 2011 [*TreeAge Software Inc 2010*]. Health economic experts were consulted for advice and guidance throughout the process of this model construction.

Figure 8.1 Diagram to represent four pain health states of a Markov model for OA.



The individual patient simulation method follows a large number of patients individually (rather than as a group) through a model where the path taken by each patient is due to

chance. That is, each patient's pathway through the model per cycle is determined by random numbers generated from a uniform distribution (0, 1). For example, using a similar model of OA with 4 pain states (no, mild, moderate and severe), assume the first patient starts in no pain health state based on a given probability; that patient's movement depends on the random number drawn from a uniform distribution ranging from 0 to 1. If the random number drawn is in the range (0 and 0.035) they move to the mild state with a probability of 0.035, if the number is between 0.035 and 0.050 they move to the moderate state with a probability of 0.015, if the number is between 0.050 and 0.055 they move to the severe state with a probability of 0.005 and if it is between 0.055 and 1 they remain in the no pain state with a probability of 0.945. Hence, the path followed by patients will differ due to random variation and thus cause the model not to be transparent as the method generally does not allow constant proportions of patients to move from one state to another per cycle. The results from this method tend to represent a sample from the population of all possible outcomes as each patient's cost and utility values can be obtained. This enables the calculation of the measure of uncertainty (variances and 95% CIs) for cost and QALYs and the probability of cost effectiveness at varied threshold values. The results of the simulation are calculated as mean costs and outcomes (QALYs) over time for the total number of patients.

The individual patient simulation method was not used, as the model developed in this study is not so complex to follow a detail patient history. However, a detailed OA model could have been built using this technique but it would require a lot of data to populate it. In addition, the technique requires high technical expertise to build it therefore it was not suitable for the timeframe or focus of this PhD.

Individual sampling model

Individual sampling models which is also a Monte Carlo simulation technique, tracks the progression of potentially different individuals (instead of cohorts) using time to next event (instead of equal cycle length). The technique facilitates the simulation of multiple attributes (e.g. status, time, age, etc) occurring in parallel for each participant in the model [*Brennan et al 2006*] and accumulates the history of each individual to determine transitions, costs, and health outcomes [*Briggs et al 2006*]. The assignment of multiple attributes overcomes the Markov assumption of homogeneity of subjects in a state without creating more states for a given model.

This technique was not used in this study because it is a complex technique which requires a lot of data to populate it and requires high technical expertise to construct it - hence it was not suitable for the timeframe of this project.

Discrete event simulations models

Discrete event simulation models describe the movement of individuals through a healthcare system which involves queuing of patients for resources, where their characteristics and outcomes are affected i.e. allowing individuals to interact with each other over unrestricted periods of time. For instance, if an individual delays in receiving an organ transplant because organs are scarce, his/her outcome is likely to affect everyone in the queue. This method was not used as the participants in this study do not interact with each other.

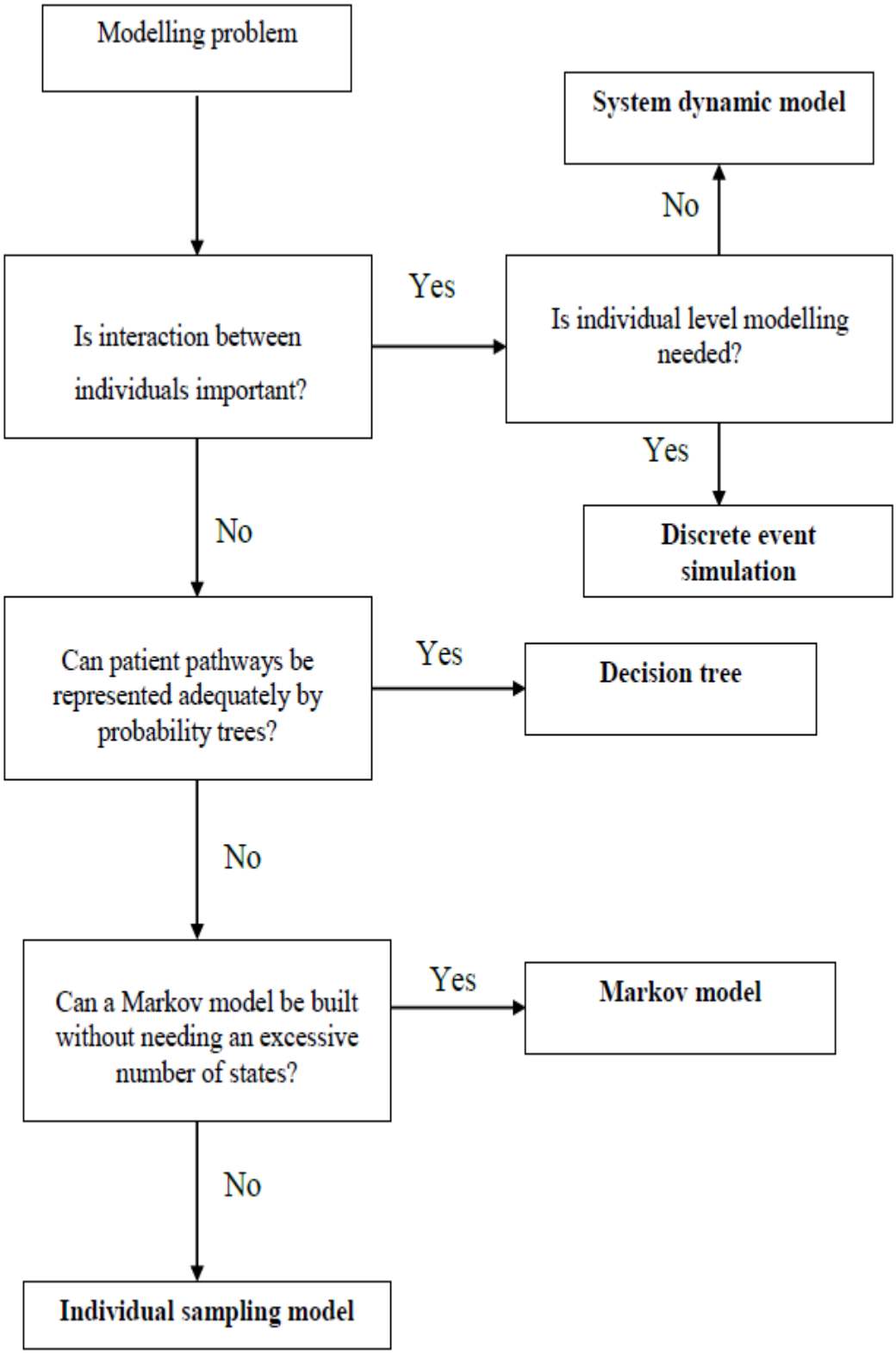
Of the above described methods, the most commonly used technique in the area of musculoskeletal research is decision tree [*Moore et al 2004; Chancellor et al 2001*] and

Markov modelling [*Maetzel et al 2001; Spiegel et al 2003*]. The individual sampling method was used by Chen et al [2006] to estimate the cost effectiveness of adalimumab, etanercept and infliximab for the treatment of rheumatoid arthritis in adults and Barton et al [2004] also used a similar method to evaluate new drugs for patients with a musculoskeletal chronic condition.

8.6.2 Choosing the appropriate decision model

Choosing the best model for a decision problem for a particular condition requires careful consideration. Barton et al [2004] carried out an overview of alternative approaches to modelling in economic evaluation and provided guidance regarding circumstances in which alternative modelling techniques should be used. Figure 8.2 below illustrates their recommendations. As can be seen, it is important to first establish whether the individuals in the model may be regarded as independent or not; where interaction between individuals is not plausible or not an important issue then the choice should be between decision tree, Markov models or individual sampling models. If interactions are likely and important, discrete event simulation or system dynamics models should be used.

Figure 8.2 Selecting an appropriate model (Barton et al 2004)



The proceeding sections present in detail the processes followed during the development of the decision model for OA in this study.

8.7 Decision model for OA

This section outlines the specifics of the model structure adopted in this study, the data and other inputs used to populate the decision model of economic evaluation that was developed to evaluate the cost effectiveness of optimal primary care interventions for OA.

8.7.1 Definition of study cohort

The relevant patient population (NorStOP cohorts) considered in this study included adults aged 50 years or more with symptomatic knee and/or hip pain (OA) in a primary care setting in the UK. The original idea was to use a similar target population as that used for the prediction modelling study (described in part 1A chapter 3 section 3.3.2 of this thesis) which comprised of participants with pain lasting three months or more at baseline at one or more joint site (hand, hip, knee or foot). The NorStOP cohort was able to provide estimates of the initial (baseline) proportion of participants among pain groups for the different joints to be included in this study. However, the proportion of participants with hand and foot OA was low (less than 5% each) in the NorStOP data, and the review (described in chapters 6 and 7 of this thesis) produced only effect estimates for hip and/or knee OA. In addition, it was possible to define no, mild, moderate and severe pain health states based on empirical data

for the WOMAC tool but not yet for the AUSCAN and FPDI tools. Therefore, it was decided in a consensus meeting that only participants with hip and knee OA should be used in this study so that the results could be appropriately linked to people with knee and/or hip OA.

8.7.2 Definition of the health states of the decision model

The model considered four health states (namely no pain, mild pain, moderate pain and severe pain) which reflect the clinical history of OA. This was a recommendation in a consensus meeting with OA clinicians, who advised that pain is widely used as the main symptom to guide treatment for OA. A health state for people who died was not considered in the model as the time horizon for the model is short and the condition is not likely to lead to a lot of deaths and secondly, it makes the model simple to implement. NorStOP participants' pain status at baseline (hip or knee) was used. The health states were defined using the baseline WOMAC scores which range from 0 to 20; participants with a score of zero were classified as no pain, 1 to 5 as mild pain, 6 to 10 as moderate pain and 11 to 20 as severe pain [*Bellamy 1996*]. However, if a participant had pain scores for both joints the highest score was used to reflect the worse joint with OA. The proportion of participants who started the model in each health state were zero for no pain health state, 0.325 for mild pain health state, 0.418 for moderate pain health state and 0.258 for severe health state. All the participants had some pain at the beginning of the study to reflect people with OA.

8.7.3 Model time horizon and cycle length

The time horizon of a model refers to the period of time over which costs and consequences of the health interventions considered in a model are measured and valued whilst time cycle refers to the duration/period over which changes in the status or symptoms of a condition is likely to occur when management options are applied. In this study, a time horizon of 3 years was adopted despite being short term for a chronic condition such as OA. It was considered appropriate as it reflects the follow up period of the NorStOP cohort adopted for this model which was used to derive the transition probabilities and the utility scores. Detailed descriptions of how the transition probabilities and utility scores were derived are presented in sections 8.7.5 and 8.7.8 respectively. A 3 -monthly time cycle was used because it was considered to be a clinically meaningful time period for OA in terms of expected changes in the symptoms of OA, duration of treatment and timing of decision making by a GP. The time horizon and cycle length were recommended in a consensus meeting with clinicians and OA experts as appropriate for OA.

8.7.4 Definition of the interventions applied

Three packages of primary care delivery were considered in this study and they are stepped care, one-stop-shop care and usual care. The first two interventions are proposed optimum care packages which were chosen because NICE has recommended that several management care options should be applied in caring for people with OA [*NICE 2008*].

Usual care was adopted from the general practitioner-led care tested in a trial by Hurley et al [*2007*] as their cost estimates were used to populate the decision model developed in this study whereas the NorStOP data provided the natural history of the people in usual care as it did not have enough data on cost of drugs used. The usual care intervention applied by Hurley et al [*2007*] reflects standard usual care for OA patients where patients are likely to be initially offered advice and pain medication(s) according to the severity and impact of their symptoms and thereafter stronger pain medications are given with surgery being the last treatment option if the pain persists, even though surgery was not considered in this study. The interventions offered by usual care include simple analgesia (paracetamol), topical NSAIDs (diclofenac gel), opioids (aspirin, codeine), NSAID (diclofenac and refecoxib tablets) and exercise.

The stepped care intervention package involves four primary care interventions namely advice and paracetamol combined, topical NSAIDs and exercise and are referred to as first, second and third line of treatment respectively (i.e. from the supposed least effective to most effective intervention). The first line of treatment (i.e. advice and paracetamol) were combined because it is uncommon in clinical practice to offer only advice if a patient has pain. The above interventions were chosen because they are recommended by NICE as the initial treatments to use in primary care for OA patients. For this package, in the beginning, all the participants are offered the first line of treatment regardless of their symptoms. As the participants condition worsens they move onto the next line of treatment up to the third line of treatment. If their condition failed to improve after the third line of treatment, they return to usual care. The movement of patients from one line of treatment to another in this study involved seeing a nurse to deliver the treatment at every step up as suggested in a consensus

meeting with clinicians and OA researchers. The exercise intervention in step 3 of the process was adopted from the exercise routine tested in a recent UK randomised trial [*Foster et al 2007*] and the associated costs over 3 months of this intervention [*Whitehurst et al 2011*]. This study's exercise cost was used because they involve adults aged 50 years or older who had been referred to 1 of 37 NHS physical therapy centres with a clinical diagnosis of knee OA, which reflects the target population of this study. Their exercise package consisted of a written advice leaflet modelled on the Arthritis Research UK leaflet on knee OA plus an individualized program which focuses on lower-limb strengthening, stretching and balancing over six sessions of 30 minutes each over a six week period. This is similar to the exercise packages used by the studies included in the review described in chapters 6 and 7.

For the one-stop-shop intervention package, participants were offered all the four primary care interventions (i.e. advice, paracetamol, topical NSAIDs, and exercise) simultaneously. This package is initially prescribed by a GP with a Physiotherapist offering the initial exercise package. The package allows all the patients to receive optimal intervention even though patients with mild symptoms may not need all the interventions. Detailed description of the application of the above package interventions is presented under the decision model structures covered in sections 8.7.6.1 to 8.7.6.3 below.

8.7.5 Transition probabilities

Transition probabilities refer to the likelihood of patients moving from one health state to another in a decision model such as a Markov model. These are usually available from clinical data and may include information such as treatment response and/or clinical event rates. The Markov model employed in this study modelled transitions between pain severity states in each 3-month time cycle over the three year duration of the model. The initial distributions in each health state at the beginning were estimated as the proportion of people with mild, moderate or severe pain (in the hip and/or knee) taken from NorStOP baseline data. Participants with no pain at baseline were excluded because the model assumed only participants with pain would consult primary care and receive treatment.

Matrix multiplication was first utilised to transform the 3 years actual transitions in the NorStOp data into 3 month transitions and further details are presented in the next section. After calculating the transition probabilities for usual care, the estimates of the standardized mean differences for each of the four primary care interventions compared to their controls (obtained from the meta-analysis described in chapters 6 and 7) were applied (multiplied) separately to the usual care transition probabilities to obtain new transition probabilities for the four primary care interventions. Table 8.1 below shows the effect estimates used for the four primary care interventions obtained from the meta-analysis carried out in Part 2A of this thesis.

Table 8.1 Effects estimates of treatments

Treatments	Effect estimate SMD (95% Confidence interval)
Advice and information	-0.17 (-0.31 to -0.03)
Simple analgesia (paracetamol)	-0.11 (-0.31 to 0.08)
Topical NSAIDs (declofenac gel)	-0.35 (-0.49 to -0.21)
Exercise	-0.32 (-0.43 to -0.21)

SMD – Standardized Mean Difference

Again, matrix multiplication was used to develop these new treatment-related transition probabilities for each health state and details of this process are also presented in a subsequent section. Even though the model began with only people with pain, it is likely that after the interventions have been applied over time, some of the participants' condition may improve and possibly move to no pain health state, hence these transition probabilities were also estimated.

8.7.5.1 Transition probabilities for usual care

For usual care, a 4 by 4 matrix was constructed in an Excel spread sheet consisting of the four health states used in this analysis. The first step was to assign arbitrary proportions to six patient transitions in the 3 months transition probability matrix. These represented the movements from each health state to the next health state. Patients could not move more than one health state higher or lower than the current health state. The probability of remaining in the same health state could be calculated by subtracting the other transitions emerging from that health state from 1.

Using matrix algebra, and assuming the 3 month transition probabilities were constant over time, a matrix showing the overall transition probabilities over a 3 year period was then calculated. In the next step, the NorStOP dataset provided the actual numbers of participants at baseline in each of the three starting health states, and these values were multiplied with each relevant row in the 3 year transition matrix (containing arbitrary values at this stage) to create a 3 year predicted number of participants matrix. Next, a matrix of the observed numbers of participants moving from one pain health state to another from baseline to 3 years (using the NorStOP data described in chapter 3) was created. The differences between the matrix with the observed numbers at 3 years and the matrix with the predicted numbers at 3 years were taken and then squared to create a matrix squared errors. These squared errors were summed, and using the solver function in Excel, these were minimized by changing the initial 6 arbitrary transitions, in order to give the matrix with the best fit (i.e. closest to the actual 3 year transitions). The spread sheet used to calculate the transition probabilities for usual care intervention can be found in appendix 10.

8.7.5.2 Transition probabilities for the four primary care interventions

To calculate the transition probabilities for the other interventions (stepped care comprising of advice/information and paracetamol combined, followed by topical NSAIDs and then exercise; and one-stop-shop comprising of all the four interventions used in stepped care combined together), copies of the usual care spread sheet containing the transition probabilities were created for each of the individual primary care interventions and one-stop-

shop. With stepped care, in each step only the effectiveness of the specific primary care intervention at that step was applied whilst with one-stop-shop care the effectiveness suggested in the consensus meeting was applied (i.e. SMD = 0.5). This estimate is arbitrary but was deemed appropriate to use as it is larger than the strongest effect estimate of the four primary care interventions.

For example, for the advice and information intervention, the 6 months predicted transition matrix of number of participants was used to estimate the overall mean and variance for the four health states used for the model. Six months was chosen arbitrarily as any time cycle could be used to arrive at the same answer. The same process was followed to calculate the overall mean and variance for the equivalent (six month) transition matrix for usual care. Copies of the advice and information spread sheet containing the above estimates were created for each of the remaining primary care interventions in the stepped care package and then for one-stop-shop intervention too. Then, standardized mean difference (SMD) was calculated between each intervention and usual care in their respective spread sheets. Next, the effect estimate (SMD) for the respective intervention obtained from the meta-analysis performed in Part 2A (also shown in Table 8.1 above) was linked using goal seek in excel to estimate the relative probability of each intervention compared to usual care. Finally, the relative probabilities values for the respective interventions were separately multiplied with the 3-months transitions probabilities of usual care to obtain the 3-months transition probabilities for the interventions respectively.

For the one-stop-shop intervention, an effect estimate of 0.5 was used as suggested by the OA experts in a consensus meeting. Also, with regards to combining advice and paracetamol

as the first line of treatment in the stepped care package, the transition probabilities of paracetamol was used as suggested by the OA experts.

A copy of the spread sheet used to calculate the transition probabilities for advice and information intervention can be found in appendix 11. Table 8.2 below shows the baseline and three monthly transition probabilities used for the respective interventions namely usual care, advice and information, paracetamol, topical NSAIDs, exercise and one-stop-shop in the decision model.

Table 8.2 Three month transition probabilities for health states for usual care, advice, paracetamol, topical NSAIDs, exercise and one-stop-shop intervention

From / To	No pain	Mild	Moderate	Severe
Baseline Probabilities	0	0.325	0.418	0.258
Usual care				
No pain	0.6022	0.3978	0	0
Mild	0.0613	0.8836	0.0551	0
Moderate	0	0.0456	0.9194	0.0350
Severe	0	0	0.0462	0.9538
Advice & Paracetamol				
No pain	0.6022	0.3978	0	0
Mild	0.1301	0.8147	0.0551	0
Moderate	0	0.0969	0.8681	0.0350
Severe	0	0	0.0980	0.9020
Topical NSAIDs				
No pain	0.6022	0.3978	0	0
Mild	0.2940	0.6509	0.0551	0
Moderate	0	0.2189	0.7461	0.0350
Severe	0	0	0.2215	0.7785
Exercise				
No pain	0.6022	0.3978	0	0
Mild	0.2729	0.6720	0.0551	0
Moderate	0	0.2032	0.7618	0.0350
Severe	0	0	0.2056	0.7944
One-stop-shop				
No pain	0.6022	0.3978	0	0
Mild	0.4010	0.5438	0.0551	0
Moderate	0	0.2986	0.6664	0.0350
Severe	0	0	0.3021	0.6979

8.7.6 Structure of the decision model

Deciding on the appropriate model structure is vital and requires a series of decisions to be made about how to relate the input parameters, including description of clinical endpoints of interest (e.g. episodes of events, disease progression, etc) that follow an underlying biological, clinical or natural history process. In this study, all the participants included in this study started with some pain (mild, moderate or severe) however, they could move into the no pain state in subsequent cycles. Transitions could occur from one health state to another when a participant's condition either improved or worsened, whilst participants whose condition remained the same stayed in the same health state. However, participants cannot move more than one health state up or down in a time cycle if the condition improves or deteriorates. The movements of the participants via the three treatment pathways are described in detail in the sections below.

8.7.6.1 Usual care sub-structure

For the usual care sub-structure in the model, during each cycle, participants in the no pain health state whose condition remains the same, stay in the same health state whilst those whose condition deteriorates move to the mild pain health state. For participants in the mild health state, those whose condition remains the same, improves or gets worse after a cycle remain in the same state (mild), move to an improved health state (no pain) or worse state (moderate) respectively. Participants in the moderate pain health state follow a similar movement as that for the mild health state where those whose condition remains the same,

improves or gets worse remain in the same state (moderate), move to an improved health state (mild) or worse state (severe) respectively. Finally, participants in the severe pain health state whose condition improve move to the moderate pain health state whereas those whose condition is unchanged remain in the severe pain health state (see Figure 8.3a below).

8.7.6.2 Stepped care sub-structure

For the stepped care intervention sub-structure, all the participants started their treatment with the first line treatment (i.e. advice and paracetamol combined). The different steps of treatment are described in section 8.7.4. The movement of the participants after a cycle into health states in this sub-structure was similar to that of the usual care sub-structure in terms of the pathways followed. However, the basic concept of this intervention's package is first consulting with a GP who also offers the first line of treatment and thereafter if participants' condition deteriorates they consult with nurse practitioner for subsequent step up treatment. Participants can only remain in the same health state or go to a health state one level worse or better; going for instance from mild to severe health state was not allowed as this is unlikely to happen over a 3 month cycle.

For the first line of treatment, for the no and mild pain health states, participants whose condition remains the same or improves stay on the same treatment, whilst those who get worse are offered the next stronger (step-up) intervention (i.e. topical NSAIDs). For participants in a moderate health state on first line treatment, an improvement in condition moves them to the mild health state and they remain on the same treatment whilst those

whose condition remains the same (moderate) or worsens stay in the same health state or move to a severe health state but both step-up to second line treatment. For participants in a severe health state on first line treatment, those whose condition improves move to a "moderate from severe" health state and remain on first line treatment as that is assumed to be a good improvement whilst those whose condition remains the same remain in the severe health state but step up their treatment (second line treatment).

For the second line of treatment, the movements of the participants into health states and their respective step-ups of interventions follows a similar pattern to that followed for the first line of treatment. However, stepping up an intervention at the second line of treatment assumes adding the third line of treatment (exercise) by consulting with a nurse practitioner to arrange the referral for 6 sessions of exercise at 30 minutes per session over a period of 6 weeks with a physiotherapist.

For the third line of treatment, movements of participants into health states were similar to those followed for the second line of treatment except that instead of stepping up treatments participants move into the respective usual care health states similar to that described in section 8.7.6.1. This is because the level of treatments ends at the third line of treatment and returning the participants to usual care will give them opportunity to received more intensive pain medications such as opioids and oral NSAIDs or onward referral for specialist care. Figure 8.3b below illustrates the structure of the first to third lines of treatments.

For the participants who moved into the "moderate from severe" health state in the respective line of treatments, those whose condition improves or remains the same move to

mild health state or stay in the same health state (moderate from severe). However, those whose condition deteriorates move into severe health state of the next line of treatment for the first and second line of treatments whereas for the third line of treatment those people move to usual care in a severe health state as there is no further step-up intervention since the model is designed for primary care patients and will also get the opportunity to receive slightly more intensive pain medications.

Those who return to the usual care follow the same pathways as described in the usual care. Figure 8.3b below includes the structure of the first to third lines of treatments for "moderate from severe" health states as well as the stepped care usual care treatments.

8.7.6.3 One-stop-shop care sub-structure

In the one-stop-shop care intervention sub-structure, the participants were offered the entire package of four primary care interventions (advice, paracetamol, topical NSAIDs and exercise) at the same time. It was assumed that as the study progresses some participants will not respond to or adhere to treatment and hence return to their GP to be placed under usual care. The movement of participants into health states in this sub-structure followed a similar pattern to that of the stepped care intervention sub-structure's third line of treatment, moderate from severe treatment and usual care treatment. The difference is that this package offers all the four primary care interventions at once to participants from the beginning whilst this occurs at the third line of treatment of the stepped care intervention.

Like the stepped care intervention, participants on this intervention who change treatment to receive usual care follow a similar structure as described in section 8.7.6.1. Similarly, for the participants who moved into the ‘moderate from severe’ health state, those whose condition remains the same, improves or worsens stay in the same health state, move to mild or severe states to receive usual care treatment respectively. The sub-structure of the one-stop-shop care intervention is illustrated in figure 8.3c below.

Figure 8.3a Markov model sub-structure for the usual care intervention

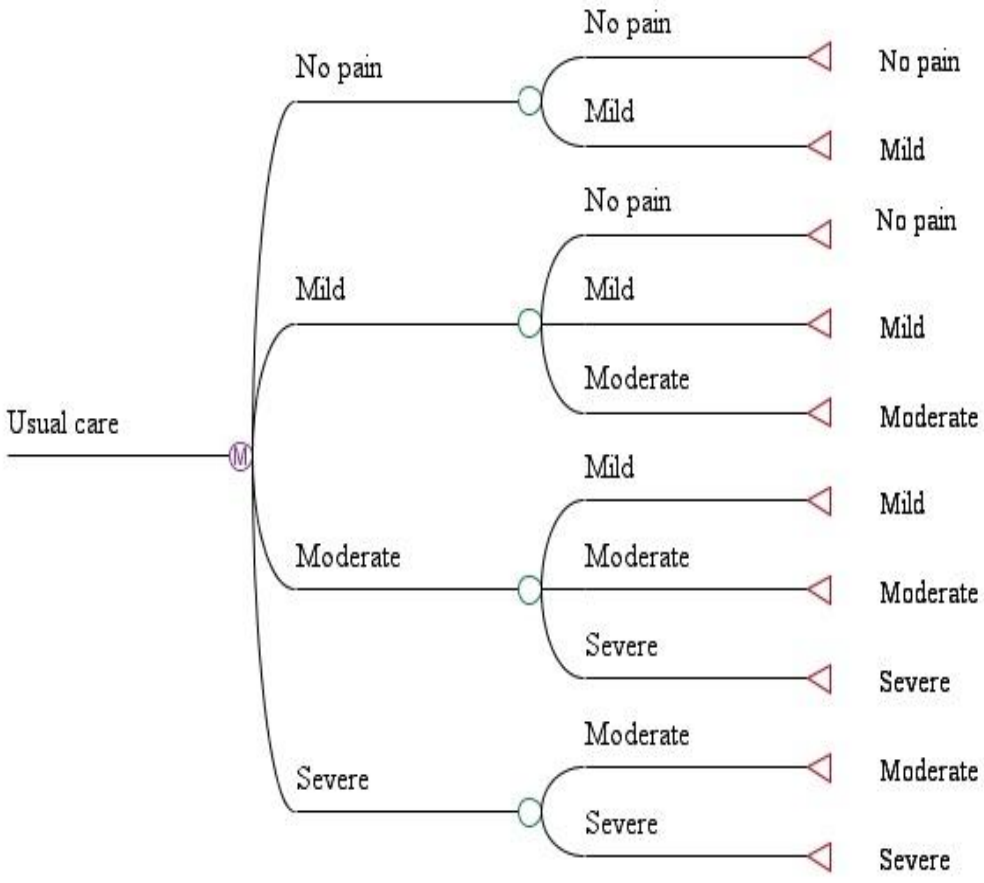


Figure 8.3b Markov model sub-structure for the stepped care intervention

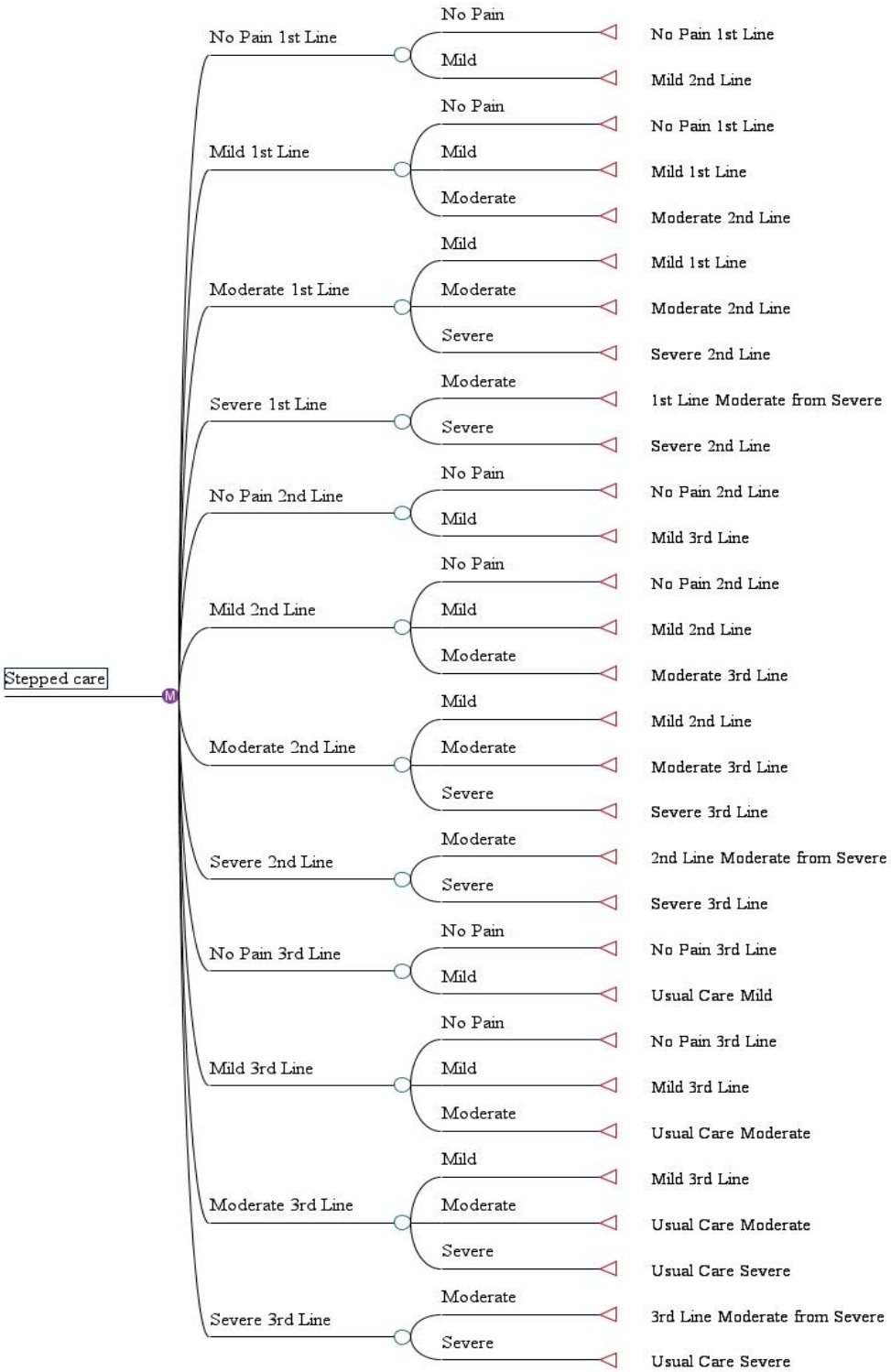


Figure 8.3b continued

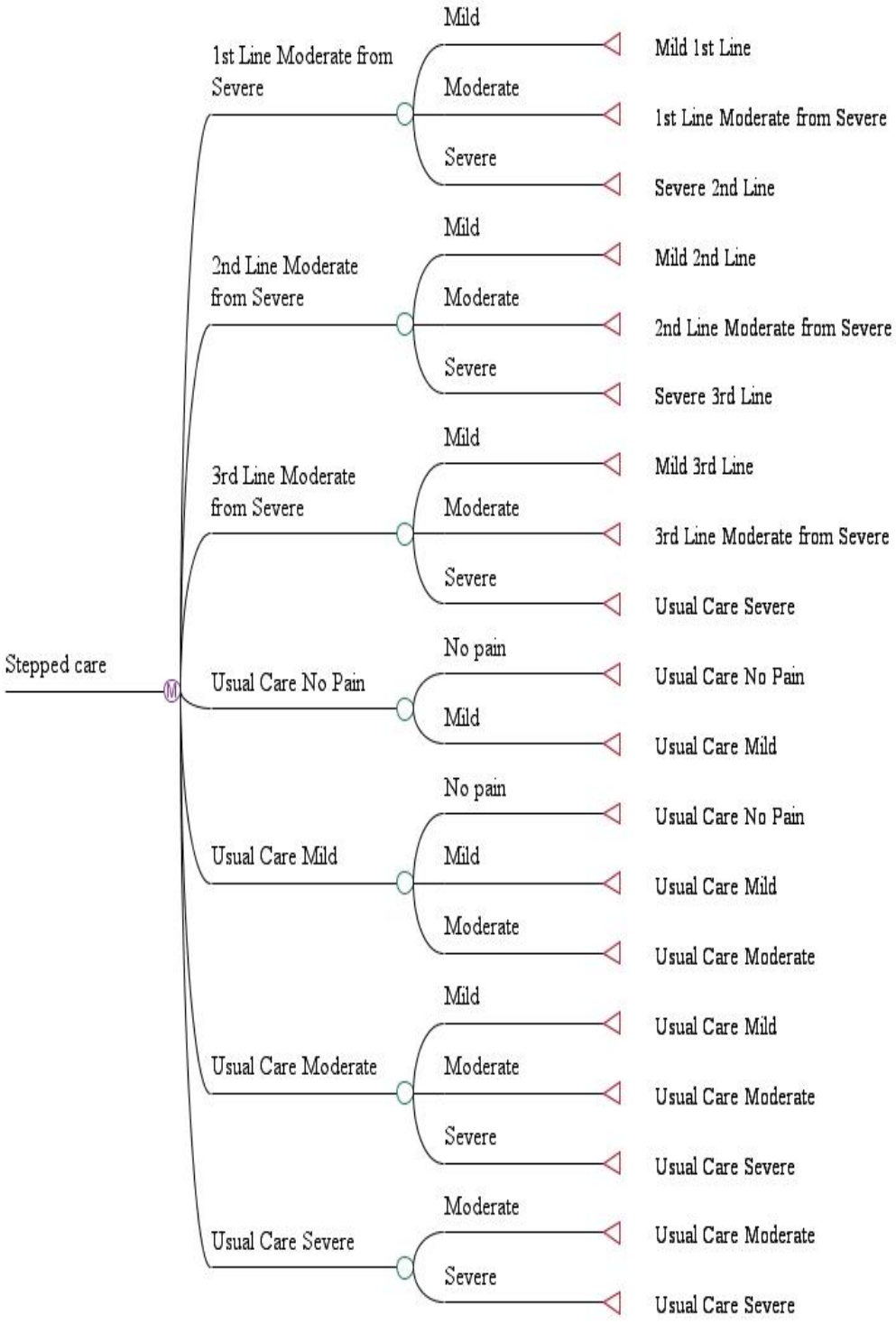
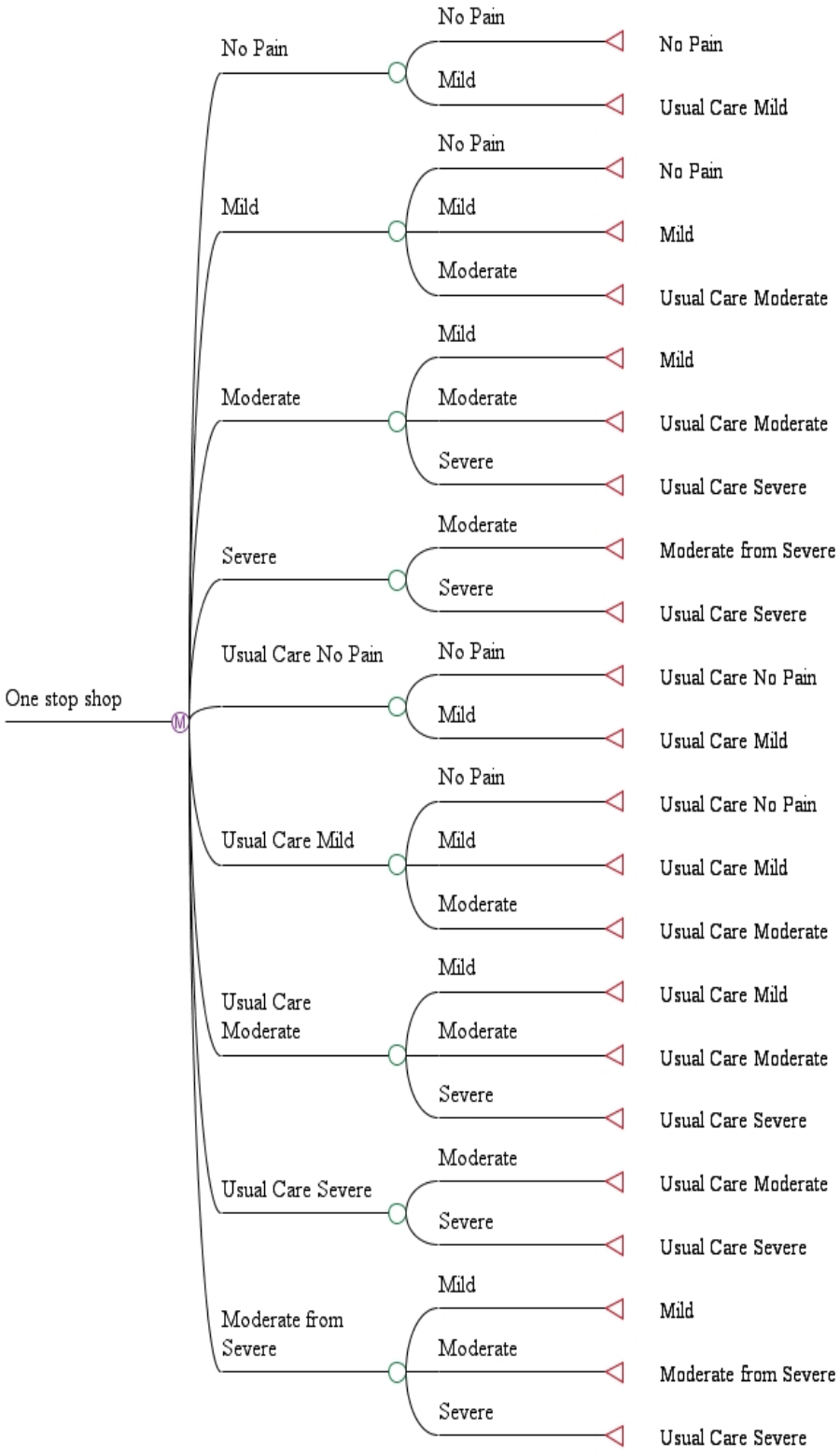


Figure 8.3c Markov model sub-structure for the one-stop-shop care intervention



8.7.7 Costs

The economic evaluation was carried out from a health care perspective which takes into account costs incurred within the health care setting. Table 8.3 below shows the unit cost data applied in this study. Two types of cost occurred in this model: one-off costs and recurring costs. The one-off costs occur when participants contact their GP for consultation at the beginning of treatment or contact with the nurse practitioner at each step-up of treatment in the stepped care package whilst the recurring costs are the repeated cost (i.e. drug cost) incurred per cycle over the three year duration of the model.

The usual care intervention cost was estimated using usual care cost data of a study carried out by Hurley et al [2007] on the clinical effectiveness of exercise, self-management and coping strategies in adults aged 50 years or more with chronic knee pain. The data comprised of six different columns of drugs some of which were paid by the NHS or purchased over the counter with their corresponding dosages provided. The number of visits to the GP was unknown (i.e. not included in the data) and as a result cost of GP consultations was not included in the cost of usual care. As the cost of the drugs within the dataset were out of date (price year 2003/4) or unknown, the unit costs (for the year 2010) of the drugs paid by the NHS were identified from the BNF and multiplied by the number of tablets taken per day to obtain the cost per day. The cost per day was multiplied by 90 days to obtain the cost for three months which was used per cycle in the decision model developed in this study. The type of drugs taken by the patients in the Hurley et al [2007] study include simple analgesia (paracetamol), topical NSAIDs (diclofenac gel), opioids (aspirin, codeine) and oral NSAIDs

(diclofenac and refecoxib tablets). For a sensitivity analysis carried out in this study, the missing data and entries of zero daily dosage were imputed using the mode number of the respective drugs and followed the process used to calculate the three months cost of drug. One way analysis of variance statistical method was used to calculate the mean usual care cost for the four health states used in this study.

For the stepped care intervention, participants incurred an initial cost for consultation with their GP. Thereafter, participants were charged for a practice nurse appointment as well as for the cost of the intervention when they move onto the next line of treatment in a new cycle. For the exercise intervention a 3 months equivalent cost for 6 sessions of exercise at 30 minutes per session over a period of 6 weeks with a physiotherapist for a year was charged. It was assumed that when participants move onto the next treatment step, all of them will continue to use the preceding intervention and hence 100% of the costs of the preceding intervention(s) were added as recommended in a consensus meeting. For example, a participant on exercise intervention would also use topical NSAIDs and paracetamol. The quarterly cost of paracetamol and topical NSAIDs were obtained from the BNF for the year 2010 whilst the cost of exercise for 3 months was taken from the study conducted by Whitehurst et al [*2011*] on the cost effectiveness of acupuncture care as an adjunct to exercise–based physical therapy for OA of the knee.

The one-stop-shop care intervention cost covers the cost of GP consultation which includes advice plus the cost of paracetamol and topical NSAIDs prescriptions plus the cost of an experienced physiotherapist (who leads the exercise intervention at an initial consultation) and cost of exercise. Subsequently, the cost of paracetamol and topical NSAIDs are charged

at each cycle. Costs were discounted at an annual rate of 3.5% in accordance with the current UK Treasury Guidelines which takes into account a time horizon more than one year over which costs are accumulated [*HM Treasury 2003*]. Costs were expressed in UK pounds sterling using 2010/2011 as the price year. A threshold of cost-effectiveness of £20,000 per QALY gained was adopted for this study [*NICE 2008a*].

Table8.3 Variables and their cost values (cost year 2010)

Individual Cost	**Cost/ person/ quarter (£)**	**Source**
Usual care cost		
No pain	9.6	Hurley et al 2007
Mild pain	24.0	Hurley et al 2007
Moderate pain	39.7	Hurley et al 2007
Severe pain	65.7	Hurley et al 2007
Paracetamol	7.78	BNF
Topical NSAIDs	16.17	BNF
Exercise	34.75	Whitehurst et al 2011
Costs of the interventions	**Unit Cost**	
GP consultation† (Same as advice)	28	PSSRU [*Curtis 2010*]
Nurse lead consultation†	14	PSSRU [*Curtis 2010*]
Physiotherapist†	34	PSSRU [*Curtis 2010*]

† - Per consultation

8.7.8 Outcomes - Quality adjusted life years (QALYs)

A QALY is the combination of quality and quantity of life and is calculated by multiplying quality of life (utility) by quantity of life (in years). In health economic modelling, the utility value for each health state is multiplied by time spent by each patient in that health state and then summed to obtain the total QALYs for that state.

In this study, the baseline SF-6D component scores of the participants in the NorStOP data (described in chapter 3) were used to derive utility values for each health state (see Table 8.4) as it was available for majority of the participants in the dataset and is also a suitable outcome for OA patients as discussed already. In order to estimate the utility values, the algorithms developed by Brazier and Roberts [2004] were applied to the SF-6D which comprised of six dimensions namely physical function, social function, role limitation, pain, mental health and vitality with 4-6 levels of response (see appendix 12 for the algorithms). The algorithm used econometric techniques to estimate health state utility values using the six dimensions of the SF-6D based on the assumption that a state with 111111 scores for the six dimensions is equal to 1 (full health) and that with death is equal to 0. However, this study does not allow or involve people who died because the time horizon of the model was short and the condition does not generally result to death. QALYs were discounted at 3.5% per annum in order to fulfil current NICE guidelines on discounting cost. Table 8.4 below shows the mean utility scores with 95% CIs estimates for the four health states used in this study.

Table8.4 Baseline mean SF-6D utility scores with 95% CI by baseline health states.

Variable	Number of subjects	Mean (95% CI)	Source
No Pain	131	0.7925 (0.7706 to 0.8144)	NorStOP WOMAC hip and knee pain score
Mild Pain	964	0.7604 (0.7518 to 0.7690)	NorStOP WOMAC hip and knee pain score
Moderate Pain	1193	0.6886 (0.6803 to 0.6969)	NorStOP WOMAC hip and knee pain score
Severe Pain	718	0.5593 (0.5499 to 0.5687)	NorStOP WOMAC hip and knee pain score
Total	3006		

8.7.9 Base case analysis

The primary/main analysis performed in an economic evaluation is known as the base case analysis. At the end of each cycle the mean cost and quality adjusted life years (QALYs) for each intervention were calculated by the model taking into consideration the events that had occurred in a cycle and the proportion in each health state. The model also calculates cumulative cost and QALYs for each intervention at each cycle and the overall mean cost and QALY values are given at the final cycle. The interventions are ordered in descending order according to cost, from least costly to most costly. Finally, the incremental cost effectiveness ratios (ICER) are calculated to compare (1) stepped care with usual care and (2) one-stop-shop care with stepped care and (3) one-stop-shop care with usual care. An ICER refers to the cost per additional QALY gain for a new intervention compared to an alternative intervention. It is calculated as the difference in costs divided by the difference in QALYs between two interventions and the value obtained is the cost per additional QALY gained for the optimal intervention compared to usual care.

8.7.10 Deterministic sensitivity analyses

This type of sensitivity analyses assumes that the input parameters are assign point estimate values and are usually varied using the lower or upper limit values. In this study, several sensitivity analyses were carried out to test the robustness of the primary results by changing some of the most important assumptions used in the model construction. The following sensitivity analyses were performed:

(i) Extension of the time horizon of the model from three years to 5, 10 and 20 years to examine if the interventions considered will continue to be effective up to 20 years which is a slightly longer period of time that may be more relevant for a chronic condition such as OA. This will help evaluate the long term effectiveness of the optimal care interventions considered in this study compared to usual care and if found to be the case will reassure both OA patients and clinicians about the long term cost-effectiveness of the optimal care interventions.

(ii) Imputation of missing data and zero scores for a particular NHS drug with the mode score of number of that drug taken per day to calculate usual care cost. By replacing missing and zero number of drugs taken per day, this will increase the cost of usual care to reflect its true value and this will help examine if this will lead to big changes in the results compared to the base case analysis.

(iii) Application of GP cost instead of nurse cost in subsequent consultations in the stepped care intervention. Both a GP led and nurse-led stepped care intervention can be used in primary care particularly in the UK though this may not be feasible in all primary care settings. Hence a sensitivity analysis to vary the lead care provider of the stepped care programme was carried out to examine changes in the results if a slightly more expensive cost of consultation (GP) is used.

(iv) 50% baseline transition probabilities for moderate and severe pain categories (instead of 0 for no pain group, 0.325 for mild group, 0.418 for moderate pain group and 0.258 for severe pain group) whilst zero values were assigned for no and mild pain categories. This will help examine changes in the findings when only people with moderate and severe OA (using equal proportion) are included in the model. The outcome of this altered model will particularly help to establish the most cost effective intervention to adopt for people with moderate to severe OA over 3 years.

(v) Application of lower and upper 95% confidence intervals of the utility scores for each health state, rather than their point estimates. This will help observe changes in the findings when utility scores are actually smaller or greater than the mean utility score used in this model.

(vi) Application of lower and upper 95% confidence intervals, rather than the point estimate, of the effect size of exercise obtained from the meta-analysis to calculate baseline transition probabilities for exercise. This will help examine the changes in the results when a lower or higher estimate of the effectiveness of exercise is used. This could have been performed for

the other primary care interventions as well – but exercise was chosen as it is an important non-pharmacological intervention for OA.

Chapter Nine

Modelling cost-effectiveness of optimal primary care for OA: Results, discussions and conclusion

This chapter outlines the results and discussion of the economic modelling study. The main results (base case analysis) are presented initially; thereafter the results of various deterministic sensitivity analyses based on the variation of the key assumptions adopted during the development of the models are presented. Finally, detailed discussions of the results and outline of strengths and limitations of the study are provided.

9.1 Base case analysis

The results of the base case analysis are presented in table 9.1 below. The table shows the treatments in order of their costs (in descending order) with their respective QALYs and cost-effectiveness. Of the three treatments, the least costly and most effective option is stepped care with one-stop-shop care being the next most effective treatment but more costly than usual care. Usual care and one-stop-shop care are described as dominated because they are more costly but less effective than stepped care – i.e. they are dominated by stepped care. As the results show a decrease in costs and a QALY gain in favour of stepped care compared to both one-stop-shop care and usual care, the ICERs estimates are not shown, as the magnitude of a negative ICER is not important.

Comparing one-stop-shop care with usual care, the former was more cost effective with an ICER of £1341 per additional QALY gained (table 9.1a).

Table 9.1 Three year costs and QALY estimates for stepped care, one-stop-shop care and usual care for adults with OA (base case).

Treatment	Cost (£)	QALYs	Difference in Cost	Difference in QALYs	ICER (Cost per QALY gained)
Stepped care	393.00	2.01	-	-	-
Usual care	427.15	1.94	34.16	-0.06	(Dominated)
One stop shop	507.62	2.00	114.62	-0.01	(Dominated)

Dominated – Treatment that costs more but is less effective than another treatment

Table 9.1a Three year costs and QALY estimates for one-stop-shop care and usual care for adults with OA (basecase).

Treatment	Cost (£)	QALYs	Difference in Cost	Difference in QALYs	ICER (Cost per QALY gained)
Usual care	427.15	1.94	-	-	-
One stop shop	507.62	2.00	80.47	0.06	1341.17

9.2 Sensitivity analyses

The first sensitivity analysis concerns the extension of the model to 5, 10 and 20 years and the results are presented in table 9.2 below showing the ranks of the treatments in descending order based on costs with their respective QALYs and ICERs where relevant. As expected both the costs and QALYs in these analyses were slightly higher for all the three treatments compared to the base case analysis with stepped care persisting to dominate both usual care and one-stop-shop care. Also, one-stop-shop care was more cost effective compared to usual care with an ICER of £747.88 (£59.83 divided by 0.08 QALYs), £311.73 (£34.29 divided by 0.11 QALYs) and £199.92 (£23.99 divided by 0.12 QALYs) for 5, 10 and 20 years respectively. This suggests that stepped care is capable of providing long term dominance up to 20 years over the other interventions by offering a higher health benefit at lower costs compared to the other interventions.

The second sensitivity analysis examines changes in the costs (and not QALYs) for the treatments when the modal number of NHS prescribed drugs taken per day for respective drugs was used for patients with missing data or zero scores to calculate usual care cost. The results are presented in table 9.3 below and as expected show slightly higher cost values but same QALY values for the three treatments as the imputed data caused the costs of the interventions to rise slightly whilst the QALYs remained fixed as they were not affected by the process compared to the base case analysis. As before, both usual care and one-stop-shop care were dominated by stepped care. One-stop-shop care continued to be more cost effective compared to usual care.

The third sensitivity analysis assumed GP rather than nurse consultations accompanied every increment in line of treatment in stepped care. Results are presented in table 9.4 below. As expected only the cost of stepped care increased by £24 (£417 - £393), with the other costs and all the QALYs remaining the same as in the base case analysis. Despite this increase, stepped care continued to dominate both one-stop-shop care and usual care.

Table 9.2 Five, ten and twenty year costs and QALY estimates for stepped care, one-stop-shop care and usual care for adults with OA.

Treatment	Costs (£)	QALYs	Difference in costs	Difference in QALYs	ICER (Cost per QALY gained)
Five year results					
Stepped care	649.94	3.24	-	-	-
Usual care	712.94	3.13	63.00	-0.11	(Dominated)
One stop shop	772.77	3.21	122.83	-0.03	(Dominated)
Ten year results					
Stepped care	1240.21	5.93	-	-	-
Usual care	1346.71	5.77	106.50	-0.16	(Dominated)
One stop shop	1381.51	5.88	141.31	-0.05	(Dominated)
Twenty year results					
Stepped care	2202.22	10.04	-	-	-
Usual care	2329.14	9.86	126.92	-0.18	(Dominated)
One stop shop	2353.13	9.98	150.90	-0.06	(Dominated)

Table 9.2a Five year costs and QALY estimates for one-stop-shop care and usual care for adults with OA.

Treatment	Costs (£)	QALYs	Difference in costs	Difference in QALYs	ICER (Cost per QALY gained)
Usual care	712.94	3.13	-	-	-
One stop shop	772.77	3.21	59.83	0.08	747.88

Table 9.3 Three year costs and QALY estimates for stepped care, one-stop-shop care and usual care using mode score of number of NHS drugs taken for respective drugs

Treatment	Costs (£)	QALYs	Difference in Cost	Difference in QALYs	ICER (Cost per QALY gained)
Stepped care	401.66	2.01	-	-	-
Usual care	444.69	1.94	43.03	-0.07	(Dominated)
One stop shop	521.35	2.00	119.69	-0.01	(Dominated)

Table 9.4 Three year costs and QALY estimates for stepped care, one-stop-shop care and usual care using GP care instead of nurse care in subsequent consultation in the stepped care arm

Treatment	Cost (£)	QALYs	Difference in Cost	Difference in QALYs	ICER (Cost per QALY gained)
Stepped care	417.25	2.01	-	-	-
Usual care	427.15	1.94	9.90	-0.07	(Dominated)
One stop shop	507.62	2.00	90.37	-0.01	(Dominated)

The fourth sensitivity analysis examines the robustness of the base case results using the 95% CI range of the utility scores for each of the four pain (no, mild, moderate and severe) health states. Figures 9a to 9d show the plots of ICERs against utility scores. The results show that for all the pain health states the ICER estimates of stepped care compared to usual care were stable across the 95% CI range of the utility scores (table 8.4) and were similar to that of the base case ICER estimates. However, for the results of stepped care compared to one-stop-shop care the ICERs increased slightly across the 95% CI range of the utility scores for no and mild pain health states whilst the ICERs appear to decrease slightly across the 95% CI range of the utility scores of the moderate and severe pain health states though they are similar compared to the base case analysis.

The fifth sensitivity analysis examines impact on the results with the assumption that only patients with at least moderate or severe pain are included in the model. Here, the baseline transition probabilities were restricted to 50% each for the two extreme pain health states (i.e. moderate and severe) and zero for the two low pain health states (i.e. no and mild). The results show slight increases and decreases in the costs and QALYs respectively for the treatments (stepped care, one-stop-shop care and usual care) compared to the base case

analysis (see table 9.5 below). The costs of all the three interventions (stepped care, usual care and one-stop-shop care) increased by £75, £95 and £70 whilst their QALYs decreased by 0.11, 0.12 and 0.11 units respectively compared to the base case analysis. This trend was observed because only participants with moderate and severe pain were included in the model and they tend to receive the most expensive interventions and have lower quality of life (utility) scores according to the structure of the model. However, stepped care continued to dominate both usual care and one-stop-shop care whilst one-stop-shop care still continued to be more cost effective compared to the usual care intervention.

The final sensitivity analysis evaluates the changes in costs and QALYs when the lower and upper 95% CI values of the effect size (SMD = 0.21 and 0.43) of exercise were used to calculate the baseline transition probabilities for exercise and used in the model (see table 9.6a and 9.6b below). This result affected only the stepped care intervention costs as expected (as it only affects the transitions in the 3rd line of treatment of stepped care) with all other estimates of costs and QALYs remaining the same for the other treatments. Compared to the base case analysis, the results after applying the lower 95% CI of the effect estimate of exercise showed a slight increase in cost (by £6) with a corresponding decrease in QALYs (by 0.01), whilst that for the upper 95% CI of the effect estimate showed a slight decrease in costs (also by £6) with no change in QALYs. For the lower effectiveness estimate, both stepped care and one-stop-shop care interventions were equally effective (with 2 QALYs) with the stepped care intervention dominating usual care and the one-stop-shop intervention being more cost effective than usual care. For the higher effectiveness estimate, stepped care dominated both usual care and one-stop-shop interventions and the results were similar to that of the base case analysis.

In summary, for the primary (base case) analysis which involved people with mild, moderate or severe pain, stepped care provided the highest health benefit followed by one-stop-shop care with usual care providing the least health benefit. In all the sensitivity analyses, one-stop-shop care continued to be more cost effective when compared with usual care intervention. In general, the changes in the costs and QALYs in the above sensitivity analyses were small but conclusions did not differ from the findings of the base case analysis, except that both stepped care and one-stop-shop care interventions resulted in equal health benefits when the lower 95% CI of the effect estimate of exercise was used to derive transition probabilities for stepped care, hence it was not possible to calculate their ICER value in that analysis but stepped care was cheaper compared to one-stop-shop care.

Figure 9a. One way sensitivity analysis of the 95% CI range for no pain health states utility score by ICER for the treatments

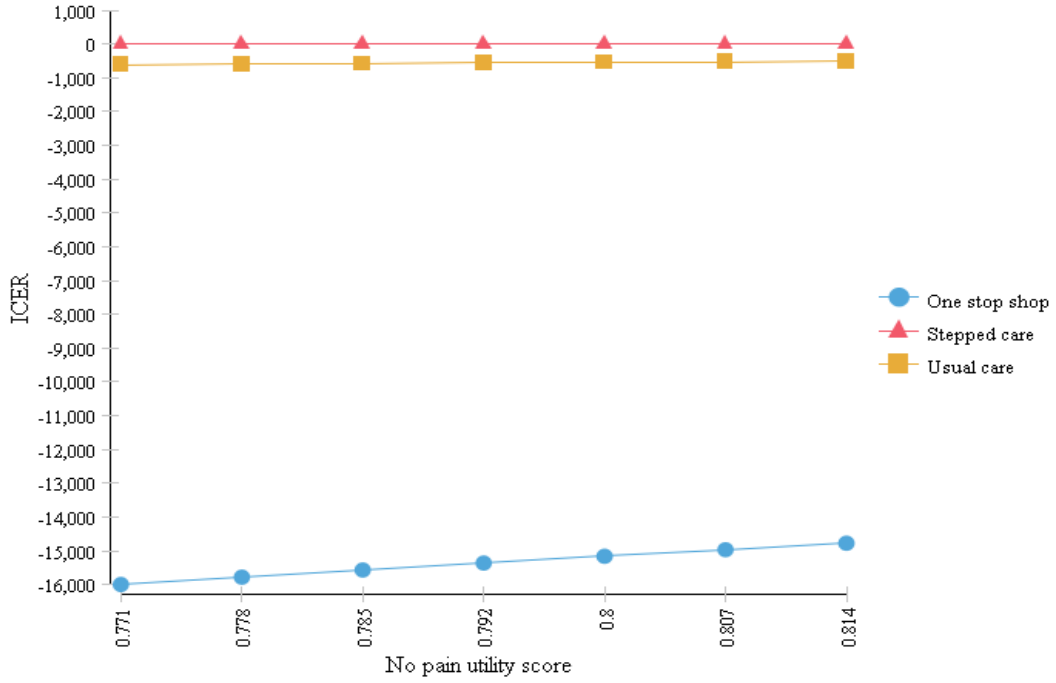


Figure 9b One way sensitivity analysis of the 95% CI range for mild pain health states utility score by ICER for the treatments

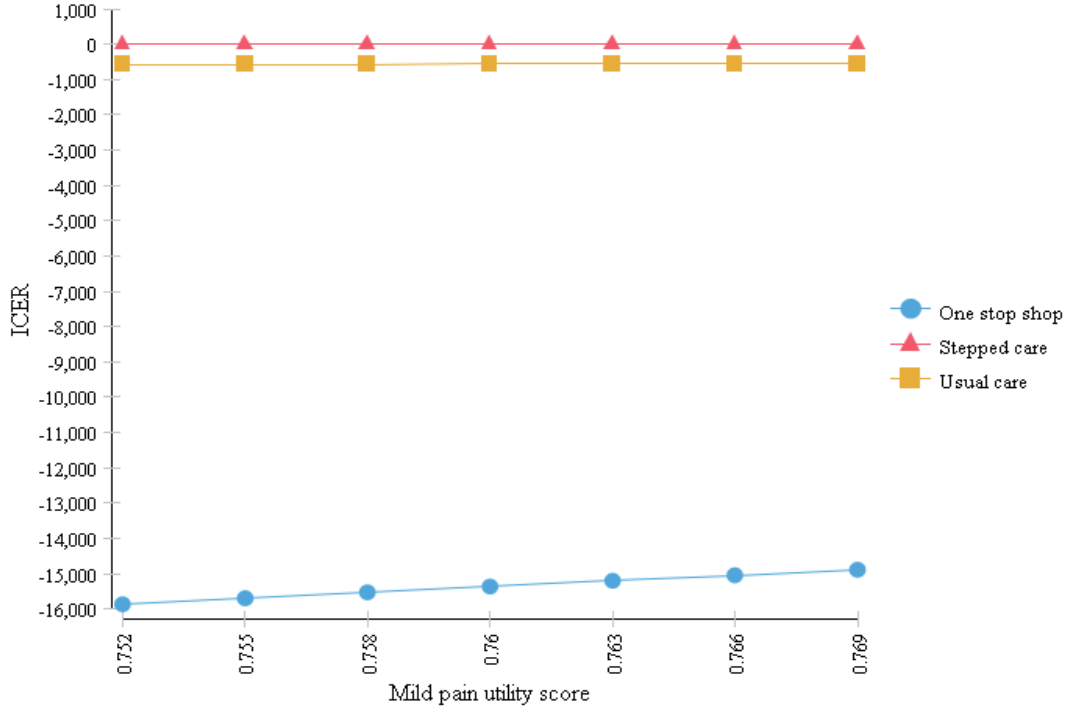


Figure 9c One way sensitivity analysis of the 95% CI range for moderate pain health states utility score by ICER for the treatments

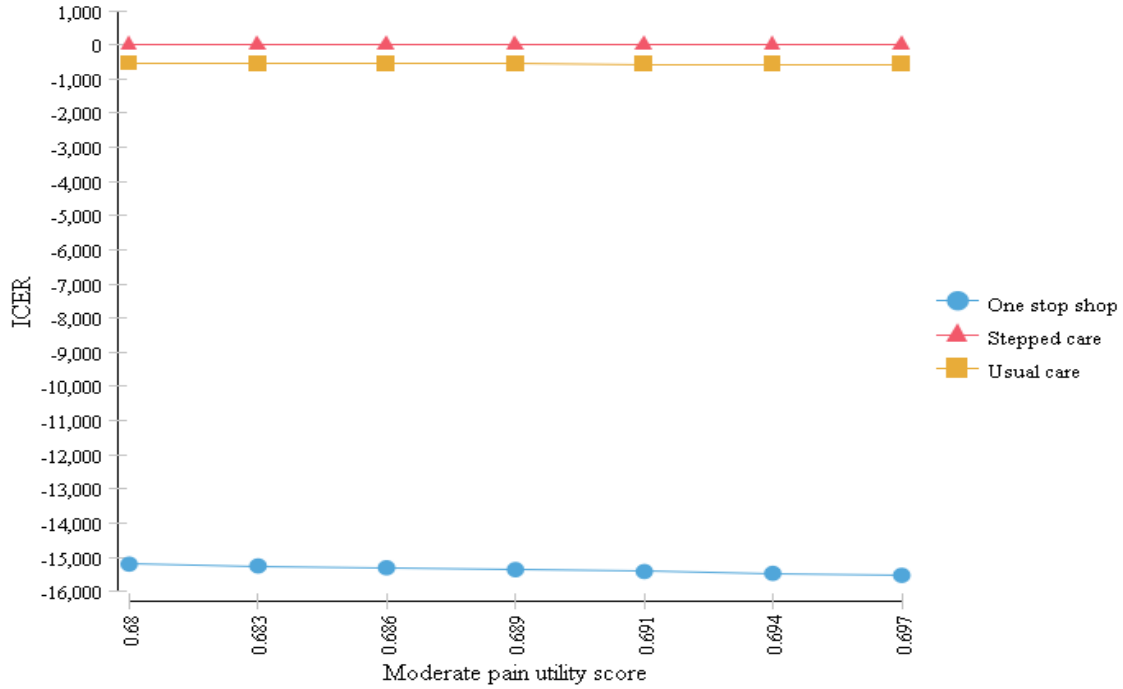


Figure 9d. One way sensitivity analysis of the 95% CI range for severe pain health states utility score by ICER for the treatments

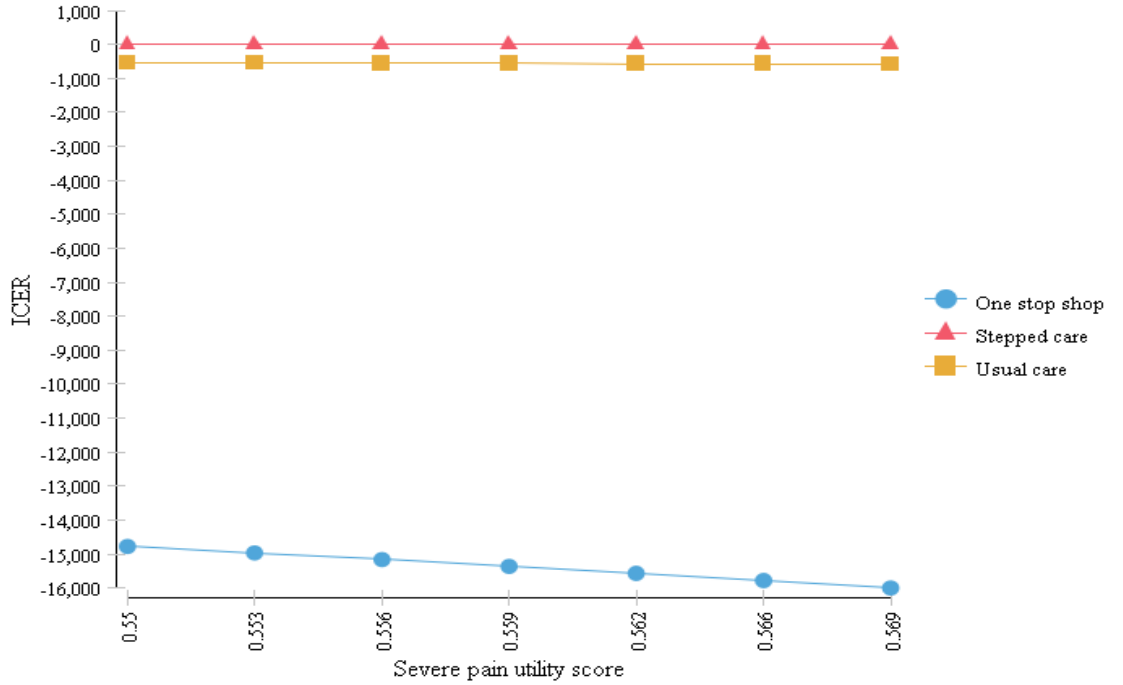


Table 9.5 Three year costs and QALY estimates for stepped care, one-stop-shop care and usual care where equal baseline transition probabilities (50 percent each) were used for moderate and severe pain health states.

Treatment	Cost (£)	QALYs	Difference in Cost	Difference in QALYs	ICER (Cost per QALY gained)
Stepped care	468.43	1.90	-	-	-
Usual care	522.26	1.82	53.83	-0.08	(Dominated)
One stop shop	577.62	1.89	109.19	-0.01	(Dominated)

Table 9.6a Three year costs and QALY estimates for stepped care, one-stop-shop care and usual care using lower 95% CI value of the effect estimate of exercise to calculate transition probabilities for exercise.

Treatment	Cost (£)	QALYs	Difference in Cost	Difference in QALYs	ICER (Cost per QALY gained)
Stepped care	399.03	2.00	-	-	-
Usual care	427.15	1.94	28.13	-0.05	(Dominated)
One stop shop	507.62	2.00	108.59	0	0

Table 9.6b Three year costs and QALY estimates for stepped care, one-stop-shop care and usual care using upper 95% CI value of the effect estimate of exercise to calculate transition probabilities for exercise.

Treatment	Cost (£)	QALYs	Difference in Cost	Difference in QALYs	ICER (Cost per QALY gained)
Stepped care	386.76	2.01	-	-	-
Usual care	427.15	1.94	40.39	-0.07	(Dominated)
One stop shop	507.62	2.00	120.86	-0.01	(Dominated)

9.3 Discussion

This section summarizes the results of the decision model presented above, describing its strengths and limitations as well as outlining general conclusions from this part of the thesis. It was not possible to compare the results with relevant findings from other published studies as there is currently no published research which has examined the cost-effectiveness of

packages of the four primary care interventions considered in this study for knee and/or hip OA.

9.3.1 Summary of findings

The results of this study have demonstrated that from the perspective of a health care system over 3 years, the stepped care intervention dominated both one-stop-shop care and usual care as it was less costly and more effective. When one-stop-shop care was compared with usual care, the former was cost-effective with an incremental cost-effectiveness ratio (ICER) of £1,300 per additional QALY gained.

The sensitivity analyses results demonstrated that the base case findings were robust to changes of these assumptions by providing similar findings in terms of costs and QALYs with the exception of the sixth assumption in which the lower 95% CI was used to calculate the transition probabilities for exercise where stepped care and one-stop-shop care interventions provided equal health benefits.

These findings of the sensitivity analyses suggest that the high health benefits provided by stepped care at a lower health care cost to adults with knee and/or hip pain extends beyond 3 years when compared to one-stop-shop care and usual care. Also, whether a GP or a nurse is consulted when patients had to move onto a stronger medication in the stepped care process makes no difference as stepped care is still cost effective. Moreover, the model can be used in clinical practice for adults aged 50 years or more with either moderate or severe pain.

However, a slight decrease in the effectiveness of exercise which may lead to fewer people having an improvement in their condition over a 3 year period of time appears to make both stepped care and one-stop-shop-care interventions equally effective.

9.3.2 Strengths and Limitations of the study

The Markov cohort transition method used to develop the decision model in this study is the most appropriate modelling technique to adopt in this context where adults aged 50 years or more with OA were able to move between different health states every 3 months for a period of 3 years to evaluate the cost effectiveness of two optimal care interventions compared to usual care. Given that the technique is best suited for chronic diseases and allows movements between disease states and the fact that the transition probabilities and utilities are based on fixed time cycles makes it convenient to appropriately represent the clinical situation/setting of the chronic condition (OA) considered in this study. This ultimately helped estimate the cost and effects of the interventions being evaluated. Moreover, the advantages associated with the cohort nature of the model are that, it makes the model easy and simple to run as well as providing a transparent movement of a cohort through the health states in a model per cycle over the duration of the model. Although a decision tree could have been used in theory, it would have required the construction of health states for each cycle over the duration of the model which would have caused the model to be extremely complex.

The main limitation of a Markov model is its memory-less assumption which states that patients' being in a particular health state is not dependent on their previous health condition

(history). This was dealt with to some extent by appropriately creating the necessary health states which reflect the general course of OA. For instance, firstly, “moderate from severe” health states were created for severe patients whose condition improved after receiving the optimal interventions and secondly, health states were created for those who moved back to usual care after their condition had not improved when receiving the third line of treatment of stepped care or one-stop-shop care packages. However, it is likely that some of the patients in a particular health state (say mild pain state) at a particular time, previously had worse symptoms (say moderate or severe pain), and this is likely to cause some level of bias in the findings as the response to treatment of patients with previous history of severe symptoms may be poorer compared to patients without a previous history of severe symptoms. Although this limitation can be resolved by creating additional health states for such groups, it would have created a model with large numbers of health states, potentially complex to implement in practice.

The NorStOP data in which adults aged 50 years or more were followed for 3 years was appropriately adopted for the decision model as it helped to estimate the transition probabilities and utility scores for those people with knee and/or hip OA. Originally, the plan was to use the same cohort (i.e. people with pain at any joint particularly knee, hip, hand or foot) as that used for the prediction model study described in chapters 3 to 6 in this thesis. However, because there was a relatively small number of people with hand or foot OA only, and because the systematic review did not produce estimates of treatment effects for hand or foot OA, and it was more convenient to use the WOMAC tool , it was decided to focus the model on knee or hip OA only. Despite this, the findings of this study would be useful to most people with OA since majority of the people who suffer the condition will have

symptoms in either their knee and/or hip. Also, since the prognosis and treatment of OA is similar irrespective of the joint affected, it is plausible to assume that the findings may also be applied to people with OA in other joints.

Although the 3-year follow up period of the NorStOP cohort of older adults reflects the 3-year time horizon considered for the decision model as well as enabled the calculation of the 3 monthly cycles considered to reflect changes in the symptoms of adults with OA, it is a disadvantage because some of the data were missing and also about a third of the participants were lost to follow up. Ideally, following participants at baseline for a short period of time, say 6 months or 1 year would have been preferred as it would have minimal loss to follow up which would have improved the estimates of the transition probabilities. Even more regular follow up of the cohort of OA patients at very short intervals say 3 monthly would have been most ideal to provide data for this model as it would have provided a better indication of the natural history of the condition (OA).

Usual care transition probabilities were calculated first, which ultimately facilitated the calculation of the transition probabilities of the four primary care interventions considered in this study by applying their effect estimates to the usual care transition probabilities to change them to reflect the effectiveness of the primary care interventions separately. The transition probabilities estimated increased the proportion of people moving from a worse health state (e.g. moderate state) to an improved health state (e.g. mild state) per cycle in the model to reflect the beneficial effect of the individual primary care treatments for OA. For the one-stop-shop care package, a stronger effect estimate (SMD = 0.5) was chosen to reflect the strength of the combined effect of the four primary care interventions offered to the

participants in the model. Although the sum of the individual effect estimates of the four primary care interventions is greater than that was used for the one-stop-shop intervention, generally in practice, when several interventions are combined and applied the effect estimate is not additive [*Foster et al 2007*] hence the estimate proposed is reasonable. The effect estimate chosen for the one-stop-shop intervention was arbitrary and even though it is bigger than the effect size of the most effective primary care intervention considered in this study, it would have been useful to vary the estimate by exploring a slightly lower and higher values of the estimate in sensitivity analyses to examine changes in the results compared to the base case. Also, it would have been useful to vary other inputs fed into the decision model such as the baseline transition probabilities, pooled effect estimates of the primary care interventions used to derive transition probabilities, time cycle, costs of health professionals consultations, etc, (using their 95% CI where possible) in sensitivity analyses to evaluated the robustness of the base case results. However, it is not possible to carry out all of these sensitivity analyses because of the time frame of this thesis and hence these could be carried out in further development of this model in future.

For the utility scores, the SF-6D components, which describe participants' preference levels of physical function, social function, role limitation, pain, mental health and vitality from the SF-36 questions administered in the same NorStOP data, were used by applying the algorithms derived by Brazier and Roberts [2004] for each of the four health states used in this study. The data used was once again appropriate in the sense that it reflects the composition of the participants used in the decision model, i.e. people aged 50 years or more with OA at various joints as this same population was used to calculate the transition probabilities. Also, the utility values calculated for this model are more likely to represent

the health states in the target population compared to utility values from the literature which may have been calculated from different patient populations. Moreover, the utility scores calculated for this model appear to be valid as they increase steadily from worse states (severe pain state) to improved state (no pain state) as expected.

The health states used for this decision model were defined using the baseline WOMAC pain scores which range from 0 to 20 where the participants with a score of zero were classified as no pain, 1 to 5 as mild pain, 6 to 10 as moderate pain and 11 to 20 as severe pain [*Bellamy 1996*]. The cut-off points used to define the groups are appropriate as they have been validated to identify levels of severity of symptoms in people with OA of the hip or knee [*Bellamy 1996*]. Even though the WOMAC tool comprises of three dimensions namely pain, stiffness and physical function for the hip or knee, the pain subscale which consist of 5 questions was used as it is one of the core outcomes for OA [*Bellamy et al 1997*] and most of the investigated primary care interventions were aimed at pain reduction. The pain subscale measures pain severity (no to severe pain) which is in line with the main outcome used in the model. Given that for several reasons only people with knee and hip OA could be considered in this model, the WOMAC tool was appropriate to use.

In this model, patients with severe pain receiving optimal care interventions and whose condition improved were assumed to move to moderate pain health state. These patients continued with the same treatment (one-stop-shop care or same line of treatment in the case of stepped care) and the results showing the trace of the cohort indicates that the proportion that moved into this assumed improved health state declines over time, as do proportions in the severe health states. As a result, the proportion of patients who return to be offered usual

care intervention increases over time which suggests that these patients end up in the usual care intervention and this may be due to the fact that most people with OA will not recover or respond sufficiently to core treatments which is a reflection of the course of OA in adults.

The 3 year follow up period considered in this study may be criticized as being too short given the prolonged nature of a chronic condition such as OA. As a result, in order to ascertain a more long term insight of the course of OA and the lasting effect of the interventions considered in this study, sensitivity analyses for 10 years and 20 years follow-up periods were investigated. The results (see table 9.2 above) demonstrated that stepped care intervention continued to provide the highest health benefits at the least health care cost followed by one-stop-shop care and usual care with the health care cost of one-stop-shop care intervention being slightly higher than that of usual care but within the threshold willingness to pay value of £20,000 recommended by NICE [*NICE 2008a*]. However, since the model followed up the participants for a long term period such as 20 years, the model would have been better represented if mortality was incorporated to reflect a natural progression of older cohort over long period. Another extension would be to include interventions such as joint replacement surgery reflecting management of OA in longer term. Yet, these issues can be investigated in future studies.

The effect estimates of the primary care interventions used in this study may not remain the same over a longer period of time and this may affect the transition probabilities where slightly fewer patients may move from a worse state to a slightly better state as time elapses. Moreover, as the population used in this model comprised of older adults, over time, they are likely to suffer with other co-morbidities which could have been incorporated in the model

although in most disease specific decision models co-morbidities and not usually incorporated in a model. As the above realities were not incorporated in the model, care must be taken when interpreting the results of this study.

The usual care cost estimate used in this study was obtained from a dataset from the Hurley et al [2007] study investigating the clinical effectiveness of integrated exercise, self-management and coping strategies for chronic knee pain. The data comprised of participants aged 50 years or more with knee pain for duration of 6 months or more recruited from primary care practices, a setting similar in many aspects to the NorStOP study, and hence it was deemed appropriate and applicable for use in this decision model. The cost of usual care was mainly based on the cost of drugs paid by the NHS since the NHS cost perspective was used for this study. The number of drugs taken per day was mostly available whilst some of the actual cost of the drugs, which relates to the year 2004, was missing. Hence the 2010 unit cost for the drugs were obtained from the BNF and applied to obtain the most recent cost. Information on participants' cost of consultation with the GP was not adequate and as a result was not added as part of usual care cost which may have under-estimated the usual care cost in this study. Even so, the usual care cost was still generally higher than stepped care cost but not one-stop-shop care. This appears to suggest that usual care potentially would have been the most costly intervention but the least effective if other costs such as GP consultations were included in the estimation of its cost. This therefore makes it plausible that optimal care for community dwelling adults with OA should be the preferred option of care as it leads to greater health improvements at a cheaper health care cost.

The design and features of the optimal care interventions employed in the decision model were suggested by clinicians and OA researchers and reflect the process of care recommended by guideline organizations such as NICE [*NICE 2004*], who advise that a holistic and optimal (combination of interventions) approach of health care should be used when managing OA. However, this study only intended to analyse the effectiveness of different ways (timing) of offering core primary care interventions. Moreover, the designs of the optimal interventions were tailored in such a way that they could be easily applied in clinical practice as a few researchers [*Smink et al 2011; The Arthritis Society 2004*] have been able to implement optimal care in clinical practice. This can be done by providing training to the clinicians (GPs and nurses) by explaining the evidence of the high improvements in health that such optimal interventions can provide to people suffering with OA. This will ultimately increase their understanding and boost their confidence to implement the optimal health care interventions [*Better Management of patients with osteoarthritis – www.boaregistret.se*] particularly if they are incorporated into health policy for implementation. However, the drawbacks of implementing optimal care are the high cost of training many or all clinicians, lack of willingness to implement new interventions on the part of some clinicians and non-adherence on the part of some patients to embrace such new interventions [*Smink et al 2011*]. Moreover, GPs and OA clinicians can argue that the current procedure of care that they apply for the management of OA generally follows either a stepped care or optimal care approaches (one-stop-shop care) – but what this study has been able to confirm is that these ways of delivering care may be highly cost-effective compared with usual care.

The fact that both optimal care packages were found to provide greater health benefits compared to usual care offers re-assurance to patients and clinicians. Even though the stepped care intervention emerged as the most attractive option followed by one-stop-shop care, the differences in their effectiveness was minimal. This suggests that potentially, a one-stop-shop care package could also be effective particularly amongst severe patient groups given that stepped care intervention commences with simple advice and paracetamol, a treatment which may not be adequate for patients suffering with severe pain. However, when the treatments in the optimal care packages cease to provide any improvements in patients' condition, the model assumed that they start usual care interventions where they get the opportunity to be offered suitably stronger medications such as opioids or referrals for surgery even though this model did not include surgery. This issue may be regarded as a limitation to this study particularly when participants are followed for longer time horizons. Also, the assumption that patients may return to usual care after 3 months may be too short as it is likely that when they stay on the same (old) intervention a little bit longer their condition might improve slightly.

To some extent the comparable benefits provided by stepped care and one-stop-shop care interventions may not be surprising given that both interventions comprised and applied all the four primary care interventions used in this study. The advantages of the stepped care package is that it will help offer patients with the right amount of treatment and the fact that nurse practitioners are used leads to decreasing the cost of consultation. However, the package may not improve the conditions of severe patients immediately as it may take longer before they are offered the right combination of treatment that could improve their condition.

The one-stop-shop package though could lead to greater improvement particularly in severe OA patients as they would be targeted with all the core primary care interventions, it may lead to waste of resources when applied to patients with mild symptoms as they may not need all the treatments in the package at one go. This is likely to make the package more costly compared to the stepped care package. However, in this model (in the one-stop-shop package) participants were asked to perform their exercise after the initial session with a physiotherapist at home where no cost is incurred, thereby reducing the cost of the package. Given the structure of the optimal care packages, it would be ideal to offer patients with severe pain the one-stop-shop package and those with low or mild pain the stepped care package. The two optimal interventions were considered in this study compared to usual care because currently, there is no evidence as to which of them is the most cost effective for managing OA in adults.

For decision modelling studies that examine the cost effectiveness of interventions, it is recommended that sensitivity analysis should take into account the uncertainty in all model parameters simultaneously, using a technique known as probabilistic sensitivity analysis (PSA) [*NICE 2004*]. This can provide information on the probability that an intervention is cost-effective at any given cost-effectiveness threshold. This type of analysis was not undertaken in this study because of the fixed time required to complete this thesis, however the model could be further developed to incorporate PSA in future. In PSA, a suitable probability distribution is defined for a model's input parameter (such as cost or utility) and samples are drawn at random from the supposed distribution to generate a single measure of the parameter which is used to estimate a measure of cost-effectiveness. The results from

PSA can help health care systems to make appropriate decisions regarding the most cost effective option to adopt for a group of people with a particular condition.

9.4 Conclusion

In summary, compared to usual care intervention, stepped care and one-stop-shop care interventions provided significant health benefits at a reasonable cost. The stepped care intervention, when compared to the one-stop-shop care intervention provided greater health benefits at a lower health care cost considering the perspective of the policymakers (NICE) economic monetary threshold, making it the most attractive intervention and potentially the option of choice in the medium to long term for patients, clinicians and healthcare systems.

Chapter Ten

Summary of findings, general discussion and conclusions

10.1 Summary of findings

This thesis involves three studies with the first study aimed to develop and validate prediction models to identify high risk predictors of OA; the second and third studies aimed to calculate the overall effect estimates of four primary care interventions for OA and subsequently feed these estimates into the decision model of which the aim was to evaluate the cost effectiveness of delivering primary care for OA.

This is the first study to develop prediction models investigating predictors of poor long-term outcome of OA regardless of the joints involved (hip, knee, hand or foot) in a population-based sample of older people. The most important predictors of pain at 3 year follow-up comprised of having knee pain at baseline, having poor physical function (SF-36) at baseline, having hand pain at baseline, not attending full time education after school and obesity whilst those that predict functional limitation at 3 year follow-up included poor physical function (SF-36) at baseline, being retired from work, performing limited activities at baseline and inability to walk for short distances (2 miles or more). The statistical models employed (Poisson and logistic regressions) fitted the data well. Also, the performance of the Poisson and logistic regression models was good for both pain and functional limitation outcome measures, although the performance estimates for the reduced models were lower, as expected. Finally, the models were deemed internally valid which suggests that they can be used in a population with similar characteristics as that of this study.

The evidence synthesis and meta-analyses evaluated the effectiveness of four primary care interventions (namely advice and information, simple analgesia, topical NSAIDs and exercise) for adults with OA and the estimates were used as inputs for the decision modelling study. The results of the risk of bias assessed in the individual RCTs showed that the majority of the RCTs were of low risk of bias, in particular those investigating topical NSAIDs which all showed low risk of bias for each of the domains assessed. Of the four interventions, advice and information (4 RCTs), topical NSAIDs (4 RCTs) and exercise (31 RCTs) demonstrated significant reduction in pain and improvement in function compared to their controls whilst simple analgesia (2 RCTs) did not significantly improve pain or functional disability. The effectiveness of the four interventions in an increasing order revealed that simple analgesia was the least effective intervention [SMD = -0.11: 95% CI -0.31 to -0.08] followed by advice/information [SMD = -0.17: 95% CI -0.31 to -0.03] and then exercise [SMD = -0.32: 95% CI -0.43 to -0.21] and topical NSAIDs [SMD = -0.35: 95% CI -0.49 to -0.21] with similar effect estimates indicating moderate effects. Study results showed no significant heterogeneity of study results. The risk of publication bias could only be assessed for the exercise intervention as it involved a sufficiently large number of RCTs with the results indicating that publication bias was unlikely.

The decision model examined the cost effectiveness of two different approaches to delivering core primary care interventions (i.e. stepped care and one-stop-shop care) compared to current primary care in older adults with knee and/or hip OA as the provision of core primary care interventions has been recommended for OA management. The findings showed that stepped care was the most attractive intervention as it provided the greatest health benefit at the lowest health care cost followed by one-stop-shop care and then current

primary care. The findings were found to be robust to changes in assumptions regarding costs, QoL scores, transition probabilities and duration of follow up with the exception of obtaining equal health benefits for stepped care and one-stop-shop care interventions when a lower effect estimate for exercise (lower 95% confidence limit instead of the point estimate), was used to generate relevant transition probabilities.

10.2 General discussion

This study is novel in the sense that it investigated optimum predictors of OA (with respect to pain and functional limitation at 3 year follow-up) involving one or more joints (i.e. hand, hip, knee or foot) as previous studies have focused on specific joints (mostly the hip or knee). Also, this study estimated PARs and NNTs for the (increasing risk) predictors identified by the models which facilitated the selection of high risk predictors of poor outcome of OA which is also rare in the literature. The evidence synthesis and meta-analyses study subsequently carried out also aimed to examine adults with OA at either of the four joint sites covered by the prediction modelling study, but this was not possible as the trials eligible for the review again only involved the hip and knee. However, the pooled estimates from the meta-analyses were fed into the decision model study which now also focused on primary care adults with hip and/or knee OA. Other data required to populate the decision model study were also mainly based on results from adults with hip and/or knee OA (in particular costs of current primary care and utility scores). The expectation is that, the findings of the decision model study may be applied to adults with similar characteristics as the prediction modelling study, as the majority of participants reported either hip or knee problems.

The combined NorStOP dataset which is comprised of adults with joint pain (OA) at baseline aged 50 years or more was used to derive the prediction models to identify the combination of factors that best predicts pain and functional limitation at 3 year follow up. The NorStOP study is a large longitudinal population-based cohort rich with data on potentially important predictors of long-term pain and disability in people with joint pain. It also contained validated self-reported sets of questions or tools such as HADS, SF-36, WOMAC, etc. However, the NorStOP datasets did not provide all the potentially relevant data to develop optimal prediction models. For instance, there was no information on radiographic measures such as changes in osteophytes and joint space narrowing which have been found in other studies [*Yusuf et al 2011; Dougados et al 1996; Ledingham et al 1993a*] to be strong predictors of OA progression in older people.

The response rate to the NorStOP survey was good with minimal difference between responders and non-responders to the baseline and 3 year follow up surveys in relation to their socio-demographic, pain and functional disability characteristics. Because of this, it is not likely that non-response at baseline and 3 years will have had a big effect on the findings of the prediction models and decision model. Muller [2010] compared the NorStOP cohort to the entire population of England and found that the NorStOP cohort had slightly higher proportion of females and lower socio-economic class compared to the English population even though the comparison was not straightforward because the sampling frame (adults registered with general practices) of the NorStOP cohort was a little different in relation to participants' age and socio-economic status. As a result, it is not easy to establish clearly if the NorStOP cohort is representative of the English population or not, which makes it

difficult to ascertain the generalizability of the findings of both the prediction and decision modelling studies.

The findings of the prediction and decision modelling studies may be applicable to community based older adults with knee and/or hip OA, but it is more difficult to assess generalizability to primary care consulters with OA, which ideally would be the target population for these models. Ideally data from primary care consulters would have been used to derive the prediction models, define OA health states, calculate utility values and baseline transition probabilities, increasing applicability and usefulness of the results of the prediction and decision model to primary care decision making. The NorStOP data however showed that few participants (9%) consulted primary care regarding their joint problems [*Jordan et al 2007*]. Hence, external validation of these models in a sample of consulters of OA will be required in future in order to establish the generalizability of the findings.

47% of the baseline variables had missing data (average of 3% per variable). When those variables with missing data of 3% or more were imputed and the analysis was repeated based on imputed datasets, the results did not change the composition of the top six strong predictors of pain and functional limitations in the prediction modeling study. The proportion of missing observations per variable in this study was low and was similar to that generally obtained for such population-based surveys [*Etter and Perneger 1997*]. Since the amount of missing data realized in the NorStOP data did not lead to a major change in the findings of the prediction models, it is also not likely to have affected the baseline transition probabilities fed into the decision model, especially as there were no missing data for the baseline WOMAC scores and minimal missing data (averaged 1%) were observed for the

SF-6D variables used to derive the utility scores for the health states used in the decision model.

The definition of OA in the prediction modelling study was decided in a consensus meeting with clinicians and OA researchers and was informed by the NICE [*NICE 2008*] and EULAR [*Zhang et al 2005 and 2007a*] OA guidelines and by availability of data from the Norstop cohorts. The resulting target population involved people with pain of the hip, knee, hand or foot for 3 months or more at baseline in the previous year, which is in line with clinical/symptomatic criteria used for OA. This criterion was applied to the NorStOP data which was used to develop and validate the prediction models for OA. Later, the same data (NorStOP survey) was used to classify people into OA health states, derive utility scores and baseline transition probabilities for the decision model which makes these estimates appropriate for use for the same target population of people meeting clinical criteria of symptomatic OA. It might have been preferable to also use radiographic data to confirm the diagnosis of OA, although X-rays are not needed to start or underpin primary care management of OA. However, it would not have been feasible to use radiography in a large population based cohort study such as NorStOP.

In the evidence synthesis and meta-analysis, pooled effect estimates of four primary care interventions (namely advice, paracetamol, topical NSAIDs and exercise) were calculated. It was considered clinically meaningful to use these core treatments to formulate "optimal" care packages for OA in a stepped care and one-stop-shop care fashions because OA is usually managed by several treatment options, and because these interventions are considered first line treatment options (core treatments) for managing OA particularly in

primary care by several OA professional organizations including NICE [NICE 2008], OARSI [*Zhang et al 2008*] and EULAR [*Pendleton et al 2000; Zhang et al 2007a*].

The effect estimates of the four primary care interventions produced by the meta-analysis study were used to derive their respective transition probabilities which were fed into the decision model to evaluate the cost effectiveness of stepped care and one-stop-shop care packages compared to usual care. However, the main shortcomings of the meta-analysis study are that very few primary care studies with generally small sample sizes were identified for these interventions (with the exception of exercise interventions), and the validity of the pooled effect estimates for the target population is not clear. It was not clear what the settings of some of the studies where especially those carried out in the USA where the health system is different from that of the UK. Moreover, the duration of treatment was on average short (compared to the cycle length of three months adopted for the decision model) which may not be ideal to realize the optimum effect of the interventions. Hence, the pooled effect estimates of these interventions used to derive transition probabilities for the decision model may not reflect the true estimates. Finally, the effect estimate (SMD=0.5) used for the one-stop-shop intervention was arbitrarily suggested by the study team, given absence of evidence from the literature on such a combined package of care. This estimate was deemed appropriate as it is larger than the effect size realized by the strongest intervention among the four primary care interventions considered in this study, yet did not assume added independent benefits of advice, medication, and exercise. Ideally, estimates from large and long follow up trials such as the BEEP (Benefits of Effective Exercise for knee Pain) and MOSAIC (Managament for OSteoArthritis In Consultation study) trials currently underway at the Arthritis Research UK Primary Care Centre can be used to

improve the decision model as they are investigating the long term effectiveness of exercise for knee pain and a model of consultation to deliver optimal primary care for OA respectively.

The pooled effect estimates obtained from the meta-analysis for the four primary interventions were fed into the decision model via deriving transition probabilities. Only the assumption regarding the effect estimate for exercise was varied in a sensitivity analysis. It would have been useful to also vary the effect estimates of the other interventions to derive transition probabilities for the decision model in order to examine the robustness of the base case findings more extensively. However, because of the time frame of this thesis and the large number of other assumptions examined, it was not feasible to perform all of these sensitivity analyses including sensitivity analyses of varying other inputs fed in the decision model such as the length of the time cycle, baseline transition probabilities, and costs of GPs and other health professionals' consultations. As a result, it is not easy to speculate whether the above variations if performed could lead to major changes in the findings compared to the base case findings.

The outcome measures considered in the prediction modelling study were pain and functional disability as they are the main consequences of OA and have been suggested as the core outcome measures for OA [*Bellamy et al 1997*]. Because of the above reasons, similar core outcome measures were considered in the evidence synthesis and meta-analysis study. For the decision modelling study, pain scores were used to create health states for OA patients. Also, the SF-6D tool was used to derive utility scores which represents QoL which was multiplied with the time spent in a specific health state (quantity of life in years) to

calculate QALYs which was used as the outcome measure for the decision model. Since the tools (WOMAC and SF-36) used to derive the above estimates are validated for use in OA patients, this will help to correctly classify people with OA as well as produce accurate QoL scores for the participants.

10.3 Implications of the studies for clinical practice and future research

The implications of the results of the prediction modelling, evidence synthesis and meta-analyses and decision modelling studies and their respective suggestions for future research are described below.

The use of RR, PAR and NNT has helped to identify the most important predictors of poor long term (3 year) outcomes of OA, which may help to identify subgroups to be targeted for early or more intensive primary care treatment. Even though other criteria/rules such as assessing only the strength or significance of the association (e.g. RR) could be used to select relevant predictors, the criteria used in this study can be considered an efficient way of selecting important high risk predictors as it combines information on strength as well as occurrence of risk indicators in the population. The predictors were selected based on large effect size (IRR), high PAR and low NNT as this is capable of selecting risk factors for which the highest feasible health benefit (IRR and PAR) and the lowest feasible effort or cost (NNT) can be if interventions would be successful. The predictors at increased risk of suffering OA in this study include those with previous knee or hand pain in previous year, poor physical function at baseline, not going onto full time education after school and being obese. Knowledge regarding predictors of increased risk of poor outcome of OA may help

GPs and other OA clinicians to target these groups early with the appropriate combination of core primary interventions and may assist health care systems when making decisions about the allocation of resources for the control and management of OA.

However, the prediction modelling study did not make the final step to develop prediction rule(s) which could be useful for GPs and OA clinicians to use in practice, most importantly because the prediction models were developed using a population-based cohort instead of a sample of OA consulters. Hence, it will be useful for future studies to go a step further and develop practically applicable prediction rule(s) after validating and updating our prediction models in clinical cohorts.

The findings of the evidence synthesis and meta-analysis study are encouraging and as a result have the potential of reassuring both OA patients and health professionals that exercise that is carried out regularly (for an average of about 5 months) can reduce pain and improve physical function related to knee and/or hip OA. Similarly, advice and information and topical NSAIDs can also lead to the improvement of pain and physical function amongst patients with knee and hip OA in primary care. However, simple analgesia appears not to be effective in relieving pain and improving function among adults with knee and/or hip OA but this may be due to unavailability of more relevant studies to establish a more accurate effect estimate for it. The findings of the meta-analysis, and hence the decision model, must be interpreted with care given that few trials were identified that concerned the setting and target population of interest for this thesis with the exception of exercise. Furthermore, very little data is available as yet to support effect estimates for the one-stop-shop approach investigated in the decision model. This implies that GPs and other OA clinicians may not be

able to directly apply the findings to their consulting patients although it could be apply to adults in the general population with knee and/or hip OA.

The results of the decision modelling study appear to provide encouragement to OA patients, health professionals and healthcare organizations that stepwise delivery of core primary care interventions may provide greater health benefits at a lower health care cost in the medium to long term compared to current usual care and to the alternative of a one-stop-shop intervention. Given that this model involved community based adults with knee and/or hip OA which are the most commonly affected joints and the fact that the management of OA is similar irrespective of the joints affected, it is plausible to assume that the model could be applicable to community based adults with OA at other joints.

Given that no one has attempted to model these types of interventions in OA before, makes it a novel model. However, the model is a first step towards designing an OA decision model which can be built upon and improved with better data in the future to enable its application in OA consulters in clinical practice.

Several recommendations for further research emerged from this study and they are as described below.

It would be useful to perform external validation to assess the performance of the prediction models in other settings and populations in future work since external validation is required to assess generalizability of a model and can be used to support the introduction of a prediction model in clinical practice. Ideally, the external validation must be performed in a

new, prospective sample of OA consulters, as this will enable the evaluation and application of the models in clinical practice.

Additional adequately powered and high quality RCTs are needed in future to evaluate the effectiveness of the four core interventions in primary care consulters with knee and/or hip OA as well as other joints including the hand and foot. This can help provide better effect estimates of the interventions that could be used to further derive inputs such as transition probabilities for the decision model which could further lead to the production of more reliable and valid estimates of the cost-effectiveness based on the decision model. The decision model would also be improved by inputs derived from a cohort of OA primary care patients with regular follow up and data collection on quality of life and resource use.

Furthermore, it would be helpful for future trials to investigate the effectiveness of optimal care packages comprising of the core primary care interventions considered in this study (compared to usual current care) over a reasonably long period of time (3 months of more) for adults with generalized OA (i.e. including hand, hip, knee or foot). This will help provide efficient estimates for the decision model for the one-stop-shop intervention since an arbitrary estimate was used in the decision model.

PSA is recommended by NICE [*NICE 2004*] as a requirement for decision modeling - however, the analysis was not undertaken in this study. It would therefore be useful to undertake this analysis in the future as it is capable of providing information on the probability that an option is cost-effective at any given willingness-to-pay threshold. The findings from PSA are robust in the sense that they tend to reflect the uncertainty around

input parameters which are used to estimate the cost effectiveness of healthcare options, which is why guideline organizations such as NICE have made it a requirement for decision models. Also, PSA findings can provide a decision maker with information on the probability that an intervention is cost-effective if the majority of the input parameters used in a model are very uncertain.

It will be useful if future decision models can be expanded to include other types of interventions such as opioids, oral NSAIDs, as well as surgical interventions. These interventions are more expensive and are associated with a higher risk of adverse events, but are commonly offered to patients with severe and chronic symptoms. Given that OA is a chronic condition, more extensive decision models would ideally cover all main interventions offered across the life span of a patient with OA. A more complex OA model might be better built such as an individual patient simulation method which is capable of varying the assumption of patients' movements in the model to evaluate the robustness of the current findings. Because the technique allows a large number of patients to be followed, its findings represent a sample from the population of all possible outcomes where each patient's cost and utility values can be obtained. However, this technique does not allow constant proportion of the patients to move from one state to another per cycle which is a suitable assumption for a chronic condition such as OA and also takes longer to run compared to the cohort simulation method.

Finally, as this study used a cohort simulation method, it will be useful for future studies to use a more complex technique such as an individual sampling simulation where patient history can be followed and patient attributes (e.g. age, gender, risk factors) can be

incorporated. The fact that these models are not tied to cycles (and look to time to next event e.g. worsening of symptoms) might be better for a disease like OA. However, high technical ability is required to build a model of this type (programming skills) and such models require detailed and reliable data on a wide range of parameters to populate them.

10.4 Conclusion

In summary, the prediction modelling study was able to select the six most important predictors of pain and functional limitation at 3 years follow up and these could be used to identify people at high risk of poor outcome of OA. The meta-analysis study findings demonstrated that core primary care interventions such as advice and information, topical NSAIDs and exercise are capable of reducing pain and improving physical function. Using the effect estimates of the meta-analysis as inputs for the decision model, the findings of the decision model showed that delivery of core interventions are capable of providing greater health benefit at a lower health care cost compared to usual current primary care with a stepped care approach being the best approach followed by a one-stop-shop approach. However, these findings need external validation, ideally in samples of primary care consulters with OA.

References

Abbot NC, Ernst E. 1998. Publication bias: direction of outcome is less important than scientific quality. *Perfusion*; 11: 182-182.

Access Economics 2001. The Prevalence, Cost and Disease Burden of Arthritis in Australia, *Report for Arthritis Australia*.

Agresti A. 1996. An Introduction to Categorical Data Analysis. *John Wiley and Sons*, Inc
Agresti A. 2007. Building and applying logistic regression models. *An Introduction to Categorical Data Analysis*. Hoboken, New Jersey: Wiley. p. 138. ISBN 978-0-471-22618-5.

AIHW - Australian Institute of Health and Welfare - Department of Health and Ageing. 2007. A picture of osteoarthritis in Australia. Canberra: Arthritis Series Number 5; AIHW (Catalogue No. PHE93). http://www.aihw.gov.au/publications/phe/apooia/apooia.pdf [Accessed on 19-08-09]

Akaike H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19(6): 716–723.

Akarca US. 2005, Gastrointestine effects of selective and non-selective non-steroidal anti-inflammatory drugs. *Curr Pharm Des*, 11:1779-93

Alexander RA, Scozzaro MJ, Borodkin LJ. 1989. Statistical and empirical examination of the Chi-Square Test for homogeneity of correlations in meta-analysis. *Psychological Bulletin*, 160; 329-331.

Altman DG and Royston P. 2006. The cost of dichotomising continuous variables. *BMJ*. 332:1080.

Altman DG, Deeks JJ. 2000. Comments on the paper by Hutton. *J. Royal Stat Soc A.* 163:415 – 416.

Altman R, Alarcon G, Appelrouth D, Bloch D, Borenstein D, Brandt K, et al. 1991 The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. *Arthritis Rheum* 34:505-14

Altman R, Alarcon G, Appelrouth D, Bloch D, Borenstein K, Brandt K, et al. 1990 The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hand. *Arthritis Rheum.*33:1601-10

Altman R, Asch E, Bloch D, Bole D, Borenstein K, Brandt K, et al. 1986. Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. *Arthritis Rheum* 29:1039-49

Altman RD, Hochberg MC, Moskowitz RW, Schnitzer TJ. 2000. Recommendations for the medical management of osteoarthritis of the hip and knee. *Arthritis Rheum* 43: 1905-1915.

Arthritis Care. OA Nation. 2004; www.arthritiscare.org.uk/OANation [Accessed on 07-08-09]

Aspden RM , Scheven BAA, Hutchison JD. 2001. Osteoarthritis as a systemic disorder including stromal cell differentiation and lipid metabolism. *The Lancet,* 357:1118 - 1120

Austin PC, Brunner LJ. 2004. Inflation of the type I error rate when a continuous confounding variable is categorized in logistic regression analyses. *Stats In Med.* 23: 1159 – 1178.

Badley EM, Tennant A. 1992. Changing profile of joint disorders with age: findings from a postal survey of the population of Calderdale,West Yorkshire, United Kingdom. *Ann Rheum Dis.* 51:366–71

Baker KR, Nelson ME, Felson DT, Layne JE, Sarno R, Roubenoff R. 2001. The efficacy of home based progressive strength training in older adults with knee osteoarthritis: a randomized controlled trial. Journal of Rheumatology; 28:1655–65.

Barton P, Bryan S, Robinson S. 2004. Modelling in the economic evaluation of health care: selecting the appropriate approach. *Journal of Health Services Research and Policy.* 9(2): 110 – 118.

Barton P, Jobanputra P, Wilson J, Bryan S, Burls A. 2004The use of modelling to evaluate new drugs for patients with a chronic condition: the case of antibodies against tumour necrosis factor in rheumatoid arthritis. *HTA.* 8(11):1366-5278

Bautch JC, Malone DG, Vailas AC. 1997. Effects of exercise on knee joints with osteoarthritis: a pilot study of biologic markers. Arthritis Care and Research;10:48–55.

Bedson J, Jordan K, Croft P. 2005. The prevalence and history of knee osteoarthritis in general practice: a case-control study. *Family practice.* 22:103-108.

Bellamy N, Campbell J, Haraoui B, Buchbinder R, Hobby K, Roth JH, et al. 2002. Dimensionality and clinical importance of pain and disability in hand osteoarthritis: development of the Australian/ Canadian (AUSCAN) Osteoarthritis Hand Index. *Osteoarthritis Cartilage.* 10:855–62.

Bellamy N. 1996. WOMAC osteoarthritis index. A users guide. London, ON: University of Western Ontario.

Belza B, Topolski T, Kinne S, Patrick DL, Ramsey SD. 2002. Does adherence make a difference? Results from a community-based aquatic exercise program. Nursing Research.

51(5): 285-291.

Bennell KL, Himan RS, Metcalf BR, Buchbinder R, McConnell J, McColl G, Green S, Crossley KM. 2005. Efficacy of physiotherapy management of knee joint osteoarthritis: a randomised, double blind, placebo controlled trial. Annals of the Rheumatic Diseases. 64:906–12.

Berkman LF, Syme SL. 1979. Social networks, host resistance, and mortality: a nine-year follow-up study of Alameda County residents. *Am J Epidemiol*, 109:186-204

Birrell F, Lunt M, Macfarlane G, Silman A. 2005. Association between pain in the hip region and radiographic changes of osteoarthritis: results from a population-based study. *Rheumatology*. 44:337–341

Biswal S, Medhi B, Pandhi P. 2006. Longterm efficacy of topical nonsteroidal antiinflammatory drugs in knee osteoarthritis: metaanalysis of randomized placebo controlled clinical trials. *J Rheumatol.* 33(9):1841-4.

BOA – Better Management of patients with osteoarthritis. http://www.boaregistret.se/en/SearchResult.aspx?IndexCatalogue=SiteIndex&SearchQuery=optimal+care Accessed 05-07-2012

Bodenheimer T. 2003. Innovations in primary care in the United States. *BMJ.* 326:796–9

Bookman AM, Williams K, Shainhouse JZ. 2004. Effect of a topical diclofenac solution for relieving symptoms of primary osteoarthritis of the knee: a randomized controlled trial. Can Med Asso Jour ;171:333-38.

Bortkewitsch Lv. 1898. Dac Gesetz der Kleinen Zahlen. *Teubner,* Leipzig.

Bowling A. 2002. Research methods in health. Investigating health and health services. *2nd edittion Buckingham*: Open University Press.

Brazier JE, Harper R, Jones NMB, O'Cathain A, Thomas KJ, Usherwood T, Westlake L. 1992. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *BMJ.* 305(6846): 160–164

Brazier JE, Roberts J, Deverill M. 2002. The estimation of a preference-based measure health from the SF-36. *J. Health Econ.* 21:271 – 292.

Brazier JE, Roberts J. 2004. The estimation of preference based measure of health from the SF-12. *Med Care.*42 (9): 851 – 859.

Brier GW. 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.*, 78, 1–3.

Briggs A, Claxton K, Sculpher M. 2006. Decision modelling for health economics evaluations. *Oxford:* Oxford University Press, Inc, New York.

Briggs AH, O'Brien BJ. 2001. The death of cost-minimization analysis? Health Economics 10(2):pp. 179-184.

British national Formulary 2010. https://www.medicinescomplete.com/mc/bnf/current/login.htm

Brooks PM. 2002. Impact of osteoarthritis on individuals and society: how much disability? Social consequences and health economic implications. *Current Opinion in Rheumatology*. 14(5): 573-577.

Brooks R. 1996. EuroQol: the current state of play. *Health policy*. 37:57 – 72.

Buckwalter JA, Saltzman C, Brown T. 2004, The impact of osteoarthritis: implications for research. *Clin Orthop Relat Res.* (427 Suppl):S6-15.

Carr AJ. 1999. Beyond disability: measuring the social and personal consequences of osteoarthritis. *Osteoarthritis Cartilage*. 7(2):230-8.

Case JP, Baliunas AJ, Block JA. 2003. Lack of efficacy of acetaminophen in treating symptomatic knee osteoarthritis. Arch Intern Med; 163:169–178.

Chernoff H, Lehmann EL. 1954. "The use of maximum likelihood estimates in chi-square tests for goodness-of-fit". *The Annals of Mathematical Statistics* 25 (3): 579–586.

Claessens AAMC, Schouten JSAG, van den Ouweland FA, Valkenburg HA. 1990. Do clinical findings associate with radiographic osteoarthritis of the knee? *Ann Rheum Dis.*49:771–4.

Cochran WG. 1954. Some methods for strengthening the common $\chi 2$ tests. Biometrics, 10; 417 – 451.

Cochrane T, Davey RC, Matthes Edwards SM. 2005. Randomised controlled trial of the cost-effectiveness of water-based therapy for lower limb osteoarthritis. Health Technology Assessmen. 9(31): iii-xi, 1-114.

Cohen J. 1977. Statistical Power Analysis for the Behavioural Sciences. New York: Academic Press.

Colditz GA, Miller JN, Mosteller F. 1989. How study design affects outcomes in comparisons of therapy. I. Medical Stat Med. 8:411-454.

Combe B, Landewe R, Lukas C, Bolosiu HD, Breedveld F, Dougados M, Emery P, et al. 2007. EULAR recommendations for the management of early arthritis: report of a task force of the European Standing Committee for International Clinical Studies Including Therapeutics (ESCISIT). *Ann Rheum Dis.* 66:34-45 doi:10.1136/ard.2005.044354 .

Cook CE, Cleland J, Pietrobon R, Garrow AP, Macfarlane GJ. 2007. Calibration of an item pool for assessing the disability associated with foot pain: an application of item response theory to the Manchester Foot Pain and Disability Index. *Physiotherapy*. 93:89-95.

Corner JL, Kirkwood CW. 1991. Decision analysis applications in the operations research literature. 1970 to 1989. *Operations Research.* 39(2): 206-219.

Creamer P, Lethbridge-Cejku M, Hochberg MC. 2000. Factors associated with functional impairment in symptomatic knee osteoarthritis. *Rheumatology*. 39:490-496.

Croft P, Cooper C, Wickham C, Coggon D. 1990. Defining osteoarthritis of the knee for epidemiological studies. *American Journal of Epidemiology.* 132(3): 514-522.

Curtis L. 2010. Unit Costs of Health and Social Care 2010. Canterbury, United Kingdom: *Personal Social Services Research Unit,* University of Kent.

Dawson J, Linsell L, Zondervan K et al. 2005. Impact of persistent hip or knee pain on overall health status in elderly people: a longitudinal population study. *Arthritis Rheum*. 53:368–74.

Deeks JJ, Higgins JPT, Altman DG (editors). Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. http://www.cochrane-handbook.org. Accessed 04-05-2011.

Department for Work and Pensions. Analytical Services Division cited in: Arthritis Research Campaign, *Arthritis the Big Picture* 2002.

DerSimonian R, Laird N. 1986. Meta-analysis in clinical trials. Control Clin Trials. 7:177-88.

Deyle GD, Henderson NE, Matekel RL, Ryder MG, Garber MB, Allison SC. 2000. Effectiveness of manual physical therapy and exercise in osteoarthritis of the knee. Annals of Internal Medicine;132(3): 173–81.

Dillon CF, Rasch EK, Gu Q, Hirsch R. 2006. Prevalence of knee osteoarthritis in the United States: Arthritis data from the Third National Health and Nutrition Examination Survey (NHANES) *J. Rheumatol.* 33:2271-9.

Ding C, Cicuttini F, Blizzard L, Scott F, Jones G. 2007. A longitudinal study of the effect of sex and age on rate of change in knee cartilage volume in adults. *Rheumatology (Oxford).* 46:273–9.

Ding C, Cicuttini F, Scott F, Gibson M, Jones G. 2003. Sex differences in knee cartilage volume in adults: role of body and bone size, age and physical activity. *Rheumatology (Oxford)*. 42:1317–23.

Doherty M, Jones AC. 1994. Osteoarthritis. *Med Int* 22:129-36.

Dominick KL, Tara K. Dudley TK, Coffman CJ, Bosworth HB. 2005. Comparison of three comorbidity measures for predicting health service use in patients with osteoarthritis. *Arthritis Care and research.* 53:666 – 672.

Dougados M, Gueguen A, Nguyen M, Berdah L, Lequesne M, Mazieres B, et al. 1996. Radiological progression of hip osteoarthritis: definition, risk factors and correlations with clinical status. *Ann Rheum Dis.* 55:356–62.

Drummond M. McGuire A. 2001. Economic evaluation in health care – merging theory with practice. *Oxford University Press Inc.* New York. USA.

Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. 2005. Methods for the Economic Evaluation of Health Care Programmes, *3rd edition. Oxford:* Oxford University Press, Inc, New York.

Eccles M, Freemantle N, Mason J. 1998. North of England evidence based guideline development project: summary guideline for non-steroidal anti-inflammatory drugs versus basic analgesia in treating the pain of degenerative arthritis. *BMJ.* 317: 526 – 530.

Efron B, Tibshirani R. 1993a. An introduction to the bootstrap. Monographs on statistics and applied probability. *New York:* Chapman & Hall.

Efron B. 1979. Bootstrap methods: Another look at the jackknife, *The Annals of Statistics*, 7, 1-26.

Efron B. 1981. Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods, *Biometrika*, 68, 589-599.

Efron B. 1982. The jackknife, the bootstrap and other resampling plans; In Society of Industrial and Applied Mathematics *CBMS-NSF Monographs*, 38.

Efron B. 1983. Estimating the error rate of a prediction rule: improvement of cross-validation. *Journal of the American statistical Association.* 78:316-331.

Egger M, Davey SG, Schneider M, Minder C. 1997. Bias in meta-analysis detected by a simple, graphical test. BMJ. 315:629–34.

Egger M, Smith GD, Altman DG. 2001. Systematic Reviews in Health Care. Meta-analysis in Context. (2nd edition). London: BMJ Books. ISBN: 072791488X.

Elders MJ. 2000. The increasing impact of arthritis on public health. *J. Rheumatol Suppl.*60:6-8.

Emkey R, Rosenthal N, Wu SC, Jordan D, Kamin M. 2004, Efficacy and safety of tramadol/acetaminophen tablets (Ultracet) as add-on therapy for osteoarthritis pain in subjects receiving a COX-2 non steroidal anti-inflammatory drug: a multicentre randomized, double-blind placebo-controlled trial. *J Rheumatol.* 31:150-6.

Englund M, Roos EM, Lohmander LS. 2003. Impact of type of meniscal tear on radiographic and symptomatic knee osteoarthritis: a sixteen-year follow up of meniscectomy with matched controls. *Arthritis and Rheumatism*. 48(8):2178–87.

Ernst E, Resch KL. 1994. Reviewer bias: A blinded experimental study. *Journal of Laboratory and Clinical Medicine*. 124: 178-182.

Etter JF, Perneger TV. 1997. Analysis of non-response bias in a mailed health survey. *Journal of Cli Epid.* 50(10):1123-1128.

Ettinger WH, Burns R, Messier SP, Applegate W, Rejeski WJ, Morgan T, Shumaker S, Berry MJ, O'Toole M, Monu J, Craven T. 1997. A randomized trial comparing aerobic exercise and resistance exercise with a health education program in older adults with knee osteoarthritis. The Fitness Arthritis and Seniors Trial (FAST). *JAMA*. 277:25–31.

Farrington DP, Loeber R. 2000. Some benefits of dichotomization in psychiatric and criminological research. *Crim Behav and Ment Health.* 10: 100 – 122.

Felson D, Goggins J, Niu J, Zhang Y, Hunter D. 2004. The effect of body weight on progression of knee osteoarthritis is dependant on alignment. *Arthritis Rheum.* 50:3904 – 9.

Felson DT, Lawrence RC, Dieppe PA, Hirsch R, Helmick CG, Jordan JM, Kington RS, Lane NE, Nevitt MC, Zhang Y, Sowers M, McAlindon T, Spector TD, Poole AR, Yanovski SZ, Ateshian G, Sharma L, Buckwalter JA, Brandt KD, Fries JF. 2000. Osteoarthritis: new insights. Part 1: the disease and its risk factors. *Ann Intern Med*. 133: 635-646.

Felson DT, Zhang Y, Anthony JM, Naimark A, Anderson JJ. 1992. Weight loss reduces the risk for symptomatic knee osteoarthritis in women. The Framingham Study. *Ann Intern Med.* 116: 535-539.

Felson DT, Zhang Y, Hannan MT, Naimark A, Weissman BN, Aliabadi P, Levy D. 1995. The incidence and natural history of knee osteoarthritis in the elderly. The Framingham Osteoarthritis Study. *Arthritis Rheum.* 38: 1500-1505.

Felson DT, Zhang Y. 1998. An update on the epidemiology of knee and hip osteoarthritis with a view to prevention. *Arthritis Rheum*; 41: 1343-1355.

Felson DT. 1988. Epidemiology of hip and knee osteoarthritis. *Epidemiol. Rev. 10:1–28.*

Ferraccioli G, Hazes JMW, Klareskog L, Machold K, Martin-Mola E, Nielsen H, Silman A, Smolen J, Yazici H. 2007. EULAR recommendations for the management of early arthritis: report of a task force of the European Standing Committee for International Clinical Studies Including Therapeutics (ESCISIT). *Ann Rheum Dis.* 66;34-45.

Foley A, Halbert J, Hewitt T, Crotty M. 2003. Does hydrotherapy improve strength and physical function in patients with osteoarthritis--a randomised controlled trial comparing a gym based and a hydrotherapy based strengthening programme. Annals of the Rheumatic Diseases. 62(12):1162–7.

Foster NE, Thomas E, Barlas P, Hill JC, Young J, Mason E, Elaine M. Hay EM. 2007. Acupuncture as an Adjunct to Exercise Based Physiotherapy for Osteoarthritis of the Knee: Randomized Controlled Trial. *BMJ.* 335:436-448.

Fransen M, Crosbie J, Edmonds J. 2001. Physical therapy is effective for patients with osteoarthritis of the knee: a randomized controlled trial. Journal of Rheumatology 28:156–64.

Fransen M, McConnell S, Bell M. 2002. Therapeutic exercise for people with osteoarthritis of the hip or knee. A systematic review. *J. Rheumatol.* 29(8):1737-45.

Fransen M, McConnell S, Hernandez-Molina G, Reichenbach S. 2009. Exercise for osteoarthritis of the hip. Cochrane Database of Systematic Reviews, Issue 3. Art. No. CD007912. DOI: 10.1002/14651858.CD007912.

Fransen M, McConnell S, Hernandez-Molina G, Reichenbach S. 2010. Does land-based exercise reduce pain and disability associated with hip osteoarthritis? A meta-analysis of randomized controlled trials. Osteoarthritis and Cartilage. 18 (5):613-620.

Fransen M, McConnell S. 2008. Exercise for osteoarthritis of the knee. Cochrane Database of Systematic Reviews, Issue 4. Art. No.: CD004376. DOI: 10.1002/14651858.CD004376.pub2.

Fransen M, Nairn L, Winstanley J, Lam P, Edmonds J. 2007. The Physical Activity for Osteoarthritis Management (PAFORM) study. A randomised controlled clinical trial evaluating hydrotherapy and Tai Chi classes. Arthritis Care and Research ;57:407–14.

Gabriel SE, Crowson CS, Campion ME, O'Fallon WM. 1997a. Indirect and nonmedical costs among people with rheumatoid arthritis and osteoarthritis compared with nonarthritic controls. *J Rheumatol.* 24(1):43-8.

Gabriel SE, Crowson CS, Campion ME, O'Fallon WM. 1997b. Direct medical costs unique to people with arthritis. *J Rheumatol.* 24(4):719-25.

Garrow AP, Papageorgiou AC, Silman AJ, Thomas E, Jayson MI, Macfarlane GJ. 2000. Development and validation of a questionnaire to assess disabling foot pain. *Pain*. 85:107–13.

Gelber AC, Hochberg MC, Mead LA 2000. Joint injury in young adults and risk for subsequent knee and hip osteoarthritis. *Annals of Internal Medicine*. 133:312–28.

Glass GV. 1976. Primary, secondary and meta-analysis of research. Educational Researcher. 5: 3-8.

Grace D, Rogers J, Skeith K, Anderson K. 1999. Topical diclofenac versus placebo: a double blind, randomized clinical trial in patients with osteoarthritis of the knee. J. Rheumatol.26:2659-63.

Grainger R, Cicuttini FM. 2004. Medical management of osteoarthritis of the knee and hip joints. *Medical Journal of Australia.* 180 (5): 232-236.

Greenland S, Drescher K. 1993. Maximum likelihood estimation of the attributable fraction from logistic models. *Biometrics* 49: 865–872.

Greenland S, Finkle WD. 1995. A critical look at methods for handling missing covariates in epidemiologic regression analyses. *Am J Epidem.* 142:1255-64.

Greenland S. 1995. Avoiding power loss associated with categorization and ordinal scores in dose-response and trend analysis. *Epidemiology.* 6:450 – 454.

Gregor S. 2006. The nature of theory in IS. *MIS Quarterly*. 30(3): 611-642

Grobbee DE, Hoes AW. 2007. Clinical epidemiology: Principles, methods, and applications for clinical research. *Sudbury, MA.* Jones and Bartlett Publishers.

Gunther KP, Sturmer T, Sauerland S, Zeissig I, Sun Y, Kessler S, Scharf HP, Brenner H, Puhl W. 1998. Prevalence of generalized osteoarthritis in patience with advanced hip and knee osteoarthritis: The Osteoarthritis Ulm Study. *Annals of the Rheumatic Disaeses.* 57(12): 717-723.

Gupta S, Hawker GA, Laporte A, Croxford R, Coyte PC. 2005. The economic burden of disabling hip and knee osteoarthritis (OA) from the perspective of individuals living with this condition. *Rheumatology*. 44:1531–153.

Haberlein TA, Baumgartner R. 1978. Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature. *Ame Soc Rev.* 43: 447-462.

Halbert J, Crotty M, Weller D, Ahern M, Silagy C. 2001. Primary care-based physical activity programs: Effectiveness in sedentary older patients with osteoarthritis symptoms. *Arthritis Rheum.*45:228–34.

Harrell FE, Lee KL, Califf RM, Pryor DB, Rosati RA. 1984. Regression modelling strategies for improved prognostic prediction. *Stat Med* 3:143–152.

Harrell FE, Lee KL, Mark DB. 1996. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 15**:** 361–387.

Hart DJ, Doyle DV, Spector TD. 1999 Incidence and risk factors for radiographic knee osteoarthritis in middle-aged women: the Chingford Study. *Arthritis Rheum.* 42: 17-24.

Hastie T, Tibshirani R, Friedman JH. 2001. The elements of statistical learning: data mining, inference and prediction. *New York*: Springer.

Hay EM, Foster NE, Thomas E, Peat G, Phelan M, Yates HE, Blenkinsopp A, Sim J. 2006. Effectiveness of community physiotherapy and enhanced pharmacy review for knee pain in people aged over 55 presenting to primary care: pragmatic randomised trial. *BMJ*;333:995.

Hayden JA, Dunn KM, van der Windt DA, Shaw WS. 2010. What is the prognosis of back pain? *Best Practice & Research Clinical Rheumatology.* 24:167 – 179.

Hedges LV, Olkin I. 1985. Statistical methods for meta-Analysis. San Diego: Academic Press.

Heller RF, Dobson AJ, Attia J, Page J. 2002. Impact numbers: measures of risk factor impact on the whole population from case-control and cohort studies. *J Epidemiol Community Health.* 56:606-610.

Helmick C, Felson D, Lawrence R, Gabriel S, et al. 2008. Estimates of the Prevalence of Arthritis and Other Rheumatic conditions in the United States. *Arthritis & Rheumatism* 58(1), 15-25.

Heuts PHTG, Bie RD, Drietelaar M, Aretz K, Hopman-Rock M, Bastiaenen CHG, Metsemakers JFM, Weel CV, Schayck OV. 2005. Self-management in osteoarthritis of hip or knee: a randomized clinical trial in a primary healthcare setting. J Rheumatol.32;543-549.

Higgins JP, Thompson SG, Deeks JJ, Altman DG. 2003. Measuring inconsistency in meta-analyses. BMJ. 327:557-60.

Higgins JP, Thompson SG. 2002. Quantifying heterogeneity in a meta-analysis. *Stat Med*. 21;1539-58.

Higgins JPT, Altman DG, Sterne JAC (editors). Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. http://www.cochrane-handbook.org. Accessed 04-05-2011.

Hildebrandt M, Bender R, Gehrmann U, Blettner M. 2006. Calculating confidence intervals for impact numbers. *BMC Medical Research Methodology.* 6:32-40.

Hinman RS, Heywood SE, Day AR. 2007. Aquatic physical therapy for hip and knee osteoarthritis: results of a single-blind randomized controlled trial. *Phys Ther.* 87:32–43.

Hirsch R, Lethbridge-Cejku M, Hanson R, Scott WW, Reichle R, Plato CC, Tobin JD, Hochberg MC. 1998. Familial aggregation of osteoarthritis: Data from the Baltimore longitudinal study on aging. *Arthritis & Rheumatism* 41: 1227–1232.

HM Treasury. 2003. The Green Book. Appraisal and Evaluation in Central Government. *Great Britain:* The Stationery Office.

Hochberg MC, Altman RD, Brandt KD, Clark BM, Dieppe PA, Griffin MR, et al. 1995a. Guidelines for the medical management of osteoarthritis. Part I. Osteoarthritis of the hip. *Arthritis Rheum* 38:1535-40.

Hochberg MC, Altman RD, Brandt KD, Clark BM, Dieppe PA, Griffin MR, et al. 1995b. Guidelines for the medical management of osteoarthritis. Part II. Osteoarthritis of the knee. *Arthritis Rheum* 38:1541-6.

Hopman-Rock M, Westhoff M. 2000. The effects of a health educational and exercise program for older adults with osteoarthritis of the hip or knee. Journal of Rheumatology. 27:1947–54.

Hordon LD, Stewart SP, Troughton PR, Wright V, Horsman A, Smith MA. 1993. Primary generalized osteoarthritis and bone mass. *Br J Rheumatol*; 32(12):1059-1061.

Hosmer DW, Lemeshow S. 2000. Applied logistic regression. *2nd edition, New York*, NY, Wiley.

Hughes SL, Seymour RB, Campbell R, Pollak N, Huber G, Sharma L. 2004. Impact of the Fit and Strong intervention on older adults with osteoarthritis. The Gerontologist. 44:217–28.

Hunink MG, Glasziou PP, Siegel JE et al. 2001. Decision making in health and medicine. Integrating evidence and values. Cambridge University Press: Cambridge.

Hunter DJ, March L, Sambrook PN. 2002. Knee osteoarthritis: the influence of environmental factors. *Clin Exp Rheumatol*; 20: 93-100.

Hurley MV, Walsh NE, Mitchell HL, Pimm TJ, Williamson E, Jones RH, Reeves BC, Dieppe PA, Patel A. 2007. Economic evaluation of a rehabilitation program integrating exercise, self-management and active coping strategies for chronic knee pain. *Arthritis and Rheumatism.* 57(7): 1220 – 1229.

Ingvarsson T, Hägglund G, Lindberg H, Lohmander LS.2000. Assessment of primary hip osteoarthritis: comparison of radiographic methods using colon radiographs. *Ann Rheum Dis.* 59:650-653.

Irwin JR, McClelland GH. 2003. Negative consequences of dichotomizing continuous predictor variables. *Journal of Marketing Research.* 40: 366 – 371.

Jacobsen S, Sonne-Holm S, Søballe K, Gebuhr P, Lund B. 2004. Radiographic case definitions and prevalence of osteoarthrosis of the hip - A survey of 4 151 subjects in the osteoarthritis sub-study of the Copenhagen City Heart Study. *Acta Orthop Scand.* 75 (6): 713–720.

Jansen MJ, Viechtbauer W, Lenssen AF, Hendriks EJ, de Bie RA. 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(1):11-20.

Jenkins CD, Stanton AB, Niemcryk SJ, Rose RM. 1988. A scale for the estimation of sleep problems in clinical research. *J Clin Epidemiol*. 41:313 – 321.

Jinks C, Jordan K, Croft P. 2006. Disabling knee pain – another consequence of obesity: Results from a prospective cohort study. *BMC Public Health.* 6: 258 – 266.

Jinks C, Jordan KP, Blagojevic M, Croft P. 2008. Predictors of onset and progression of knee pain in adults living in the community. A prospective study. *J. Rheumatology*. 47:368-374.

Jinks C, Lewis M, Ong BN, Croft P. 2001. A brief screening tool for knee pain in primary care: Validity and reliability. *Rheumatology* (*Oxford*). 40(5): 528 – 536.

Jordan K, Clarke AM, Symmons DPM, Fleming D, Porcheret M, Kadam UT, Croft P. 2007. Measuring disease prevalence: a comparison of musculoskeletal disease using four general practice consultation databases. *British Journal of General Practice*. 57(534): 7–14.

Jordan J, Luta G, Renner J, Dragomir A, Hochberg M, Fryer J. 1997. Knee pain and knee osteoarthritis severity in self-reported task specific disability: The Johnson County Osteoarthritis Project. *J. Rheumatol*. 24:1344 – 5.

Jordan JM, Helmick CG, Renner JB, Luta G, Woodard J, Dragomir AD, Woodard J, Fang F, Schwartz TA, Nelson AE, Abbate LM, Callahan LF, Kalsbeek WD, Hochberg MC. 2009. Prevalence of hip symptoms and radiographic and symptomatic hip osteoarthritis in African-Americans and Caucasians: the Johnston County Osteoarthritis Project. *J Rheumatol*. 36:809–15.

Jordan JM, Helmick CG, Renner JB, Luta G, Woodard J, Dragomir AD, Woodard J, Fang F, Schwartz TA, Abbate LM, Callahan LF, Kalsbeek WD, Hochberg MC. 2007. Prevalence of knee symptoms and radiographic and symptomatic knee osteoarthritis in African-Americans and Caucasians: the Johnston County Osteoarthritis Project. *J Rheumatol.* 34:172–80.

Juni P, Altman DG, Egger M. 2001. Systematic reviews in health care: Assessing the quality of controlled clinical trials. *BMJ.* 323:42-6.

Jünia P, Holensteinb F, Sternea J, Bartletta C, Egger M. 2002. Direction and impact of language bias in meta-analyses of controlled trials: empirical study. *Int. J. Epidemiol*. 31 (1): 115-123.

Keefe FJ, Blumenthal J, Baucom D, Affleck G, Waugh R, Caldwell DS, Beaupre P, Kashikar-Zuck S, Wright K, Egert J, Lefebvre J. 2004. Effects of spouse-assisted coping skills training and exercise training in patients with osteoarthritic knee pain: a randomized controlled study. *Pain.* 110:539–49.

Keefe FJ, Caldwell DS, Williams DA, Gil KM, Mitchell D, Robertson C, Martinez S, Nunley J, Beckham JC, Helms M. 1990. Pain coping skills training in the management of osteoarthritic knee pain – II: Follow up results. *Behav. Therapy*. 21(435 – 447).

Kellgren JH, Lawrence JS. 1957. Radiologic assessment of osteo-arthrosis. Ann *Rheum Dis.* 16:494- 502.

Kristman V, Manno M, Cote P. 2004. Loss to follow-up in cohort studies: how much is too much? *Eur J Epidemiol.* 19(8):751-60.

Lapsley HM, March LM, Tribe KL, Cross MJ, Brooks PM. 2001, Living with osteoarthritis: patient expenditures, health status and social impact. *Arthritis Rheum.* 45:301-306.

Lau EC, Cooper C, Lam D et al. 2000. Factors associated with osteoarthritis of the hip and knee in Hong Kong Chinese: obesity, joint injury and occupational activities. *American Journal of Epidemiology.* 152:855–62.

Laupacis A, Sekar N, Stiell IG. 1997. Clinical prediction rules: a review and suggested modifications of methodological standards. *Journal of the American Medical Association.* 277:488 - 494.

Lawrence RC, Felson DT, Helmick CG, Arnold LM, Choi H, Deyo RA, Gabriel S, Hirsch R, Hochberg MC, Hunder GG, Jordan JM, Katz JN, Kremers HM, Wolfe F. 2008. National Arthritis Data Workgroup: Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part II. *Arthritis Rheum,* 58(1):26-35.

Lawrence RC, Felson DT, Helmick CG, Arnold LM, Choi H, Deyo RA, Gabriel S, Hirsch R, Hochberg MC, Hunder GG, Jordan JM, Katz JN, Kremers HM, Wolfe F. 2008. National Arthritis Data Workgroup: Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part II. *Arthritis Rheum,* 58(1):26-35.

Lawrence RC, Helmick CG, Arnett FC, Deyo RA, Felson DT, Giannini EH, Heyse SP, Hirsch R, Hochberg MC, Hunder GG, Liang MH, Pillemer SR, Steen VD, Wolfe F. 1998. Estimates of the prevalence of arthritis and selected musculoskeletal disorders in the United States. *Arthritis & Rheum.* 41: 778 – 799.

Leardini G, Salaffi F, Caporali R, Canesi B, Rovati L, Montanelli R, et al. 2004. Direct and indirect costs os oateoarthritis of the knee. *Clinical & Experimental Rheumatology.* 22(6): 699-706.

Ledingham J, Dawson S, Preston B, Milligan G, Doherty M. 1993a. Radiographic progression of hospital referred osteoarthritis of the hip. *Ann Rheum Dis*. 52:263–7.

Ledingham J, Regan M, Jones A, Doherty M. 1993 Radiographic patterns and associations of osteoarthritis of the knee in patients referred to hospital. *Ann Rheum Dis.* 52: 520-6.

Leigh JP, Seavey W, Leistikow B. 2001, Estmating the costs of job related arthritis. *J Rheumatol* 28:164-1654.

Levy E, Ferme A, Perocheau D, Bono I. 1993. Les couts socio-economiques de l'arthrose en France. *Rev. Rhum. Ed Fr.* 60:63S-67S.

Lewis SC, Langman MJ, Laporte JR, Matthews JN, Rawlins MD, Wiholm BE. 2002, Dose-response relationship between individual nonaspirin non-steroidal anti-inflammatory drugs (NANSAIDs) and serious upper gastrointestinal bleeding: a meta-analysis based on individual patient data. *Br J Clin Pharmacol.* 54:320-6.

Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, et al. 2009. The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. Ann Intern Med, 151(4):W-65–W-94.

Lin SY-C, Davey RC, Cochrane T. 2004. Community rehabilitation for older adults with osteoarthritis of the lower limb: a controlled clinical trial. Clinical Rehabilitation. 18(1): 92-101.

Little RA. 1992. Regression with missing X's; a review. *J. Am Stat Assoc.* 87:1227-37.

Lord J, Victor C, Littlejohns P, Ross FM, Axford JS. 1999. Economic evaluation of a primary care-based education programme for patients with osteoarthritis of the knee. *Health Technology Assessment.* 3(23):1-55.

Loza E, Lopez-Gomez JM, Abasolo L, Maese J, Carmona L, Batlle-Gualda E. 2009. Economic Burden of Knee and Hip Osteoarthritis in Spain. *American College of Rheumatology*, 61(2):158 – 165.

Lui KJ. 2004. Statistical estimation of epidemiological risk. *Chichester.* John Wiley & Sons Ltd.

Lybrand S. 2003. Health related quality of life and the burden of disease in Australian rheumatology practice [Masters thesis]. *Brisbane:* University of Queensland.

MacCallum RC. Zhang S, Preacher KJ, Rucker DD. 2002. On the practise of dichotomization of quantitative variables. *Psychol Methods.* 7: 19 – 40.

MacLean CH, Knight K, Paulus H, Brook RH, Shekelle PG. 1998. Costs attributable to osteoarthritis. *J Rheumatol.* 25(11):2213-8.

Maetzel A, Krahn M, Naglie G. 2001. The cost-effectiveness of celecoxib and rofecoxib in patients with osteoarthritis or rheumatoid arthritis. Technology Report No 23. Ottawa: Canadian Coordinating Office for Health Technology Assessment.

Mahoney MJ. 1977. Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research*; 1: 161-175.

Mallen CD, Peat G, Thomas, Lacey R, Croft P. 2007, Predicting poor functional outcome in community-dwelling older adults with knee pain: prognostic value of generic indicators. *Ann Rheum Disorders.* 66:1456–1461.

Manek NJ, Hart D, Spector TD, MacGregor AJ. 2003. The association of body mass index and osteoarthritis of the knee joint. *Arthritis Rheum*. 48:1024–9.

Manek NJ, Lane NE, 2000. Osteoarthritis: Current Concepts in Diagnosis and Management. *Journal of American Family Physicians*. 61:1795-804. http://www.aafp.org/afp/20000315/1795.html [Accessed 30-10-09].

Mannoni A, Briganti MP, Di Bari M, Ferrucci L, Costanzo S, Serni U, Masotti G, Marchionni N. 2003. Epidemiology profile of symptomatic osteoarthritis in older adults: A population based study in Dicomano, Italy. *Ann Rheu Dis.* 62:576-578.

March LM, Bachmeier CJ. 1997. Economics of osteoarthritis: a global perspective. *Baillieres Clin Rheumatol.* 11:817-34.

Mason CH and Perreault WD. 1991. Collinearity, Power, and Interpretation of Multiple Regression Analysis. *Journal of Marketing Research.* 28(3): 268-280.

Maurer BT, Stern AG, Kinossian B, Cook KD, Schumacher HR. 1999. Osteoarthritis of the knee: isokinetic quadriceps exercise versus an educational intervention. *Archives of Physical Medicine and Rehabilitation.* 80:1293–99.

Mazzuca SA, Brandt KD, Katz BP, Hanna MP, Melfi CA. 1999. Reduced utilization and cost of primary care clinic visits resulting from self-care education for patients with osteoarthritis of the knee. *Arthritis & Rheumatism*. 42(6):1267 – 1273.

McAlindon T, Cooper C, Kirwan J, Dieppe P. 1993. Determinants of disability in osteoarthritis of the knee. *Ann Rheum Dis.* 52:258 – 62.

McAlindon TE, Cooper C, Kirwan JR, Dieppe PA. 1992a Knee pain and disability in the community. *Br J Rheumatol* 31:189–92.

McAlindon TE, Snow S, Cooper C, Dieppe PA. 1992b Radiographic patterns of osteoarthritis of the knee joint in the community: the importance of the patellofemoral joint. *Ann Rheum Dis* 51:884–9.

McCullagh P, Nelder JA. 1989. Generalized Linear Models – Monographs on Statistics and Applied Probability 37. *Second Edition.* Chapman and Hall, London, UK.

McDuffie FC, Felts WR, Hochberg MC et al. 1987. Arthritis and musculoskeletal diseases. In: Amler RW & Dull HB (eds). Closing the gap: the burden of unnecessary illness. Oxford: *Oxford University Press*, 19–29.

Menz HB, Tiedmann A, Kwan MM, Plumb K, Lord SR. 2006. Footpain in community-dwelling older people: an evaluation of the Manchester Foot Pain and Disability Index. *Rheumatology*. 45:863-7.

Messier SP, Loeser RF, Miller GD, Morgan TM, Rejeski WJ, Sevick MA, Ettinger WH, Pahor M, Williamson JD. 2004. Exercise and dietary weight loss in overweight and obese older adults with knee osteoarthritis. *Arthritis and Rheumatism.* 50:1501–10.

Mikesky AE, Mazzuca SA, Brandt KD, Perkins SM, Damush T, Lane KA. 2006. Effects of strength training on the incidence and progression of knee osteoarthritis. *Arthritis Care and Research.* 55:690–99.

Miller ME, Langefeld CD,. Tierney WM, Hui SL, McDonald CJ. 1993. Validation of probabilistic predictions. *Med Decis Making* 13: 49–58.

Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. 2007. Epidemiology and reporting characteristics of systematic reviews. PLoS Med 4: e78. doi:10.1371/journal.pmed.0040078.

Molinaro AM, Simon R, Pfeiffer RM. 2005. Prediction error estimation: a comparison of resampling methods. *Bioinformatics.* 21(15): 3301 – 3307.

Montgomery MP, Kamel F, Hoppin JA, Beane-Freeman LE, Alavanja MCR, Sandler DP. 2010. Characteristics of non-participation and potential for selection bias in a prospective cohort study. *Am J Ind Med.* 53(5): 486–496.

Moons K, Royston P, Vergouwe Y, Grobbee D, Altman D. 2009. Prognosis and prognostic research: what, why and how? *BMJ. 23:*338:b375.

Moore A, Phillips C, Hunsche E, Pellissier J, Crespi S. 2004. Economic evaluation of etoricoxib versus non-selective NSAIDs in the treatment of osteoarthritis and rheumatoid arthritis patients in the UK. Pharmacoeconomics. 22(10):643-60.

Muller S, 2010. The measurement of locomotor disability in epidemiological studies. [*PhD Thesis*]. Keele University.

Murphy L, Schwartz TA, Helmick CG, Renner JB, Tudor G, Koch G, Dragomir A, Kalsbeek WD, Luta G, Jordan JM. 2008. Lifetime risk of symptomatic knee osteoarthritis**.** *Arthritis Rheum*, 59:1207-1213.

Murphy L, Schwartz TA, Helmick CG, Renner JB, Tudor G, Koch G, Dragomir A, Kalsbeek WD, Luta G, Jordan JM. 2008. Lifetime risk of symptomatic knee osteoarthritis**.** *Arthritis Rheum.* 59:1207-1213.

Nagelkerke NJ. 1991. A note on a general definition of the coefficient of determination. *Biometrika*. 78:691-692.

NCC-CC – National Collaborating Centre for Chronic Conditions. Osteoarthritis: National clinical guideline for care and management in adults. London: Royal College of Physicians, 2008. Accessed on 27-10-09 http://www.nice.org.uk/nicemedia/pdf/CG059FullGuideline.pdf

Neame RL, Muir K, Doherty S, Doherty M. 2004. Genetic risk of knee osteoarthritis: a sibling study. *Ann Rheum Dis*. 63:1022–1027.

Nelder, John A; Wedderburn, Robert W (1972). "Generalized linear models". *Journal of the Royal Statistical Society Series A* (Journal of the Royal Statistical Society. Series A (General), Vol. 135, No. 3) 135 (3): 370–384.

Nevitt MC, Xu L, Zhang Y, Lui LY, Yu W, Lane NE, Qin M, Hochberg MC, Cummings SR, Felson DT. 2002. Very low prevalence of hip osteoarthritis among Chinese elderly in Beijing, China, compared with whites in the United States. *Arthritis Rheum.* 46:1773-1779

NICE – National Institute for Health and Clinical Excellence. 2008a. Guide to the methods of technology appraisal.

NICE – National Institute for Health and Clinical Excellence. 2004. Guide to the methods of technology appraisal. *London: National Institute for Health and Clinical Excellence.* UK.

NICE – National Institute for Health and Clinical Excellence – Osteoarthritis, The care and management of osteoarthritis in adults, 2008 http://www.nice.org.uk/nicemedia/pdf/CG59NICEguideline.pdf

Niethard FU, Gold MS, Solomon GS, Liu JM, Unkauf M, Albrecht HH, Elkik F. 2005. Efficacy of Topical Diclofenac Diethylamine Gel in Osteoarthritis of the knee. J Rheumatol. 32: 2384 – 92.

Nuki G 1998. Diseases of the connective tissues, joints and bones. In: Edwards CRW, Bouchier IAD, Haslett C et al. (eds). Davidson's principles and practice of medicine. Edinburgh: Churchill Livingstone, 864–939.

O'Reilly SC, Muir KR, Doherty M. 1996. Screening for pain in knee osteoarthritis: which question? *Ann Rheum Dis.* 55:931–3.

O'Reilly SC, Muir KR, Doherty M. 1999. Effectiveness of home exercise on pain and disability from osteoarthritis of the knee: a randomized controlled trial. *Annals of the Rheumatic Diseases.*58:15–19.

Odding E, Valkenburg A, Algra D, Vandenouweland FA, Grobbee DE, Hofman A. 1998. Associations of radiological osteoarthritis of the hip and knee with locomotor disability in the Rotterdam Study. *Ann Rheum Dis*. 57:203–8.

Olson CM, Rennie D, Cook D, Dickersin K, Flanagin A, Hogan JW, Zhu Q, Reiling J, Pace B. 2002. Publication bias in editorial decision making. *JAMA*; 287: 2825-2828.

Patrick DL, Ramsey SD, Spencer AC, Kinne S, Belza B, Topolski TD. 2001. Economic evaluation of aquatic exercise for persons with osteoarthritis. *Medical Care.* 39(5):413–24.

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. *Ann Rheum Dis.* 60:91-97.

Peat G, Thomas E, Handy J, Wood L, Dziedzic K, Myers H, Wilkie R, Duncan R, Hay E, Hill J, Croft P. 2004. The Knee Clinical Assessment Study - CAS(K). A prospective study of knee pain and knee osteoarthritis in the general population, *BMC Musculoskeletal Disorders*, 5(4):1-9.

Peloquin L, Bravo G, Gauthier P, Lacombe G, Billiard JS. 1999. Effects of a cross-training exercise program in persons with osteoarthritis of the knee. A randomized controlled trial. *Journal of Clinical Rheumatology* 5:126–36.

Pendleton A, Arden N, Dougados M, Doherty M, Bannwarth B, Bijlsma JWJ, Cluzeau F, Cooper C, Dieppe PA, Günther KP, Hauselmann HJ, Herrero-Beaumont G, Kaklamanis PM, Leeb B, Lequesne M, Lohmander S, Mazieres B, Mola EM, Pavelka K, Serni U, Swoboda B, Verbruggen AA, Weseloh G, Zimmermann-Gorska I. 2000. EULAR recommendations for the management of knee osteoarthritis: report of a task force of the Standing Committee for International Clinical Studies Including Therapeutic Trials (ESCISIT). *Ann Rheum Dis.* 59:936-944.

Peterson IF & Jacobssen LT 2002. Osteoarthritis of the peripheral joints: best practice and research. *Clinical Rheumatology.* 16(5):741–60.

Petrou S, Gray A. 2011b. Economic evaluations using decision analytical modelling: design, conduct, analysis and reporting. *BMJ*. 342:d1766.

Pincus T, Koch G, Lei H, Mangal B, Sokka T, Moskowitz R, Wolfe F, Gibofsky A, Simon L, Zlotnick S, Fort JG. 2004. Patient preference for placebo, acetaminophen or celecoxib efficacy studies (PACES): two randomized, double blind, placebo controlled, crossover clinical trials in patients with knee or hip osteoarthritis. *Ann Rheum Dis*. 63:931–939.

Praemer AP, Furner S, Rice DP. 1999. Musculoskeletal Conditions in the United States. Rosemont, Illinois, *American Academy of Orthopaedic Surgeons* 182.

Quilty B, Tucker M, Campbell R, Dieppe P. 2003. Physiotherapy, including quadriceps exercises and patellar taping, for knee osteoarthritis with predominant patello-femoral joint involvement: randomized controlled trial. *Journal of Rheumatology*. 30: 1311–17.

Rabenda V, Manette C, Lemmens R, Mariani AM, Struvay N, Reginster JY. 2006. Direct and indirect costs attributable to osteoarthritis in active subjects. *J. Rheum.* 33:1152-8.

Ravaud P, Flipo RM, Boutron I, Roy C, Mahmoudi A, Giraudeau B. 2009. ARTIST (osteoarthritis intervention standardized) study of standardised consultation versus usual care for patients with osteoarthritis of the knee in primary care in France: pragmatic randomised controlled trial. BMJ. 338:b421.

Ravaud P, Giraudeau B, Logeart I, Larguier JS, Rolland D, Treves R, et al. 2004. Management of osteoarthritis (OA) with an unsupervised home based exercise programme and/or patient administered assessment tools: a cluster randomized controlled trial with a 2x2 factorial design. *Ann Rheum Dis*.63:703–8.

Reed LJ, Berkson J. 1929. The application of the logistic function to experimental data. *J. Physical Chemistry.* 33: 760 – 779.

Reginster JY. 2002, The prevalence and burden of arthritis. *J Rheumatology* (Oxford*)* 41(Supp 1) 3-6.

Reijman M, Hazes JMW, Pols HAP, Bernsen RMD, Koes BW, Bierma-Zeinstra SMA. 2004. Validity and reliability of three definitions of hip osteoarthritis: cross sectional and longitudinal approach. *Ann Rheum Dis*. 63:1427–1433.

Robinson R. 1993c. Economic evaluation and health care. Cost effectiveness analysis. *British Medical Journal*. 307: 793 – 795.

Robinson R. 1993e. Economic evaluation and health care. Costs benefit analysis. *British Medical Journal*. 307: 924 – 926.

Roddy E, Muller S, Thomas E. 2009. Defining disabling foot pain in older adults: further examination of the Manchester Foot Pain and Disability Index. *Rheumatology.* 48(8):992-996.

Roth SH, Shainhouse JZ. 2004. Efficacy and safety of a topical diclofenac solution (Pennsaid) in the treatment of primary osteoarthritis of the knee. A randomized, double-blind, vehicle-controlled clinical trial. *Arch Intern Med.*164:2017-23.

Rothstein HR, Sutton AJ Borenstein M. 2005. Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments. *Wiley and Sons Ltd.* Chichester, England.

Royston P, Altman DG, Sauerbrei W. 2006. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med.* 25:127 – 141.

Rubin DB. 1976. Inferences and missing data. *Biometrica.* 63:581-90.

Rubin DB. 1987. Multiple imputation for non response in surveys. *New York.* Wiley.

Rucker G, Schwarzer G, Carpenter JR, Schumacher M. 2008. Undue reliance on I^2 in assessing heterogeneity may mislead. *BMC Med Res Methodol*. 27;8:79. doi: 10.1186/1471-2288-8-79.

Sa KN, Pereira CdeM, Souza RC, Baptista AF, Lessa I. 2011. Knee Pain Prevalence and Associated Factors in a Brazilian Population Study. *Pain Medicine.* 12: 394–402pme

Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. 1996. Evidence-based medicine: what is it and what it isn't. *BMJ.* 312: 71 – 72.

Schafer JL. 1997.Analysis of incomplete multivariate data. *Chaman & Hall/CRC Press*, London.

Schellingerhout JM, Heymans MW, de Vet HC, et al. 2009. Categorizing continuous variables resulted in different predictors in a prognostic model for nonspecific neck pain. *J Clin Epidemiol.* 62:868–74.

Schouten JS, de Bie RA & Swaen G. 2002. An update on the relationship between occupational factors and osteoarthritis of the hip and knee. *Current Opinions in Rheumatology* 14(2):89–92.

Schwarz GE. 1978. Estimating the dimension of a model. *Annals of Statistics* 6(2): 461–464.

Sculpher MJ, Claxton K, Drummond M, McCabe C. 2006. Whither trial-based economic evaluation for health care decision making? *Health Economics.* 15:677-87.

Seferlis T, Lindholm L, Nemeth G. 2000. Cost-minimisation analysis of three conservative treatment programmes in 180 patients sick-listed for acute low-back pain. *Scandinavian Journal of Primary Health Care.* 18(1):53–57.

Sharma L, Cahue S, Song J, Hayes K, Pai Y, Dunlop D. 2003. Physical functioning over three years in knee osteoarthritis. Role of psychosocial, local mechanical and neuromuscular factors. *Arthritis Rheum*. 48:3359 – 370.

Sharma L. 2001. Local factors in osteoarthritis. *Curr Opin Rheumatol*. 13: 441-446.

Shmueli G, Koppius O, 2009. Predictive vs. Explanatory Modeling in IS Research. *Robert H. Smith School Research Paper No. RHS 06-058 http://ssrn.com/abstract=1112893* Accessed on 21-04-10.

Smink AJ, van den Ende CH, Vliet Vlieland TP, Swierstra BA, Kortland JH, Bijlsma JW, Voorn TB, Schers HJ, Bierma-Zeinstra SM, Dekker J. 2011. Beating osteoARThritis": development of a stepped care strategy to optimize utilization and timing of non-surgical treatment modalities for patients with hip or knee osteoarthritis. *Clin Rheumatol*. 30(12):1623-9.

Smit F, Ederveen A, Cuijpers P, Deeg D, Beekman A. 2006. Opportunities for Cost-effective Prevention of Late-Life Depression. An Epidemiological Approach. *Arch Gen Psychiatry*. 63:290-296.

Smit F, Comijs H, Schoevers R, Cuijpers P, Deep D, Beekman A. 2007. Target groups for the prevention of late life anxiety. *British Journal of Psychiatry*. 190:428 – 434.

Sowers MF. 2001. Epidemiology of risk factors for osteoarthritis: systemic factors. *Curr Opin Rheumatol.* 13: 447-451.

Spector TD, Hart DJ, Leedham-Green M. 1991. The prevalence of knee and hand osteoarthritis in the general population using different clinical criteria: the Chingford Study. *Arthritis Rheum*. 34(suppl):S171.

Spiegel BM, Targownik L, Dulai GS, Gralnek IM. 2003. The cost-effectiveness of cyclooxygenase-2 selective inhibitors in the management of chronic arthritis. *Ann Intern Med.* 138:795–806.

StataCorp. 2009. Stata Statistical Software: Release version 11. *College Station, Texas*: StataCorp LP.

Sterne JAC, Gavaghan D, Egger M. 2000. Publication and related bias in meta-analysis: Power of statistical tests and prevalence in the literature. *Journal of Clinical Epidemiology.* 53: 1119-1129.

Steyerberg EW , Eijkemans MJ , Harrell FE Jr , Habbema JDF. 2000. Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. *Stats Med.* 19:1059 -79.

Steyerberg EW, Bleeker SE, Moll HA, Grobbee DE, Moons KGM. 2003. Internal and external validation of predictive models: A simulation study of bias and precision in small samples. *J. Clin. Epidemiol*. 56: 441 -447.

Steyerberg EW, Eijkemans MJ, Habbema JD. 1999. Stepwise selection in small data sets: a simulation study of bias in logistic regression analysis. *J Clin Epidemiol* 52: 935–942.

Steyerberg EW, Harrell FE Jr, Borsboom GJJM, Eijkemans MJC, Vergouwe Y, Habbema JDF. 2001. Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. *J. Clin. Epidemiol*. 54(8): 774-781.

Steyerberg EW. 2009. Clinical prediction models – A practical approach to development, validation and updating. *Statistics for biology and health.* New York: Springer Science + Business Media, LLC.

Superio-Cabuslay E, Ward MM, Lorig KR. 1996. Patients education interventions in osteoarthritis and rheumatoid arthritis: A meta-analytic comparison with non-steroidal anti-inflammatory drug treatment. *American College of Rheumatology*. 9(4): 292 – 301.

Tak E, Staats P, Van Hespen A, Hopman-Rock A. 2005. The effects of an exercise program for older adults with osteoarthritis of the hip. *Journal of Rheumatology*. 6:1106–13.

Talbot LA, Gaines JM, Huynh TN, Metter EJ. 2003. A home-based pedometer-driven walking program to increase physical activity in older adults with osteoarthritis of the knee: a preliminary study. *Journal of the American Geriatric Society.* 51:387–92.

The Arthritis Society. 2004. Getting a grip on arthritis. Best practice guideline. http://www.acreu.ca/pdf/Best-Practice-Guidelines.pdf. Accessed 05-07-2012

Thomas E, Harris L, Wilkie R, Croft PR. 2004a, The prevalence of pain and pain interference in a general population of older adults: cross-sectional finding from the North Staffordshire Osteoarthritis Project (NorStOP). *Pain.* 110(1-2):361-8.

Thomas E, Peat G, Mallen 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? *Ann Rheum Disorder* 67:1390–1398.

Thomas E, Wilkie R, Peat G, Hill S, Dziedzic KS, Croft PR. 2004b, The North Staffordshire Osteoarthritis Project - NorStOP: Prospective, 3 year study of the epidemiology and management of clinical osteoarthritis in the general population of older adults. *BMC Musculoskelet Disord.* 1471-2474/5/2 5:2.

Thomas KS, Muir KR, Doherty M, Jones AC, O'Reilly SC, Bassey EJ. 2002. Home based exercise programme for knee pain and knee osteoarthritis: randomised controlled trial. *BMJ.* 325:752–7.

Tompa E, Culyer AJ, Dolinschi R, 2008. Economic evaluation of interventions for occupational health and safety – Developing good practice. *Oxford University Press.* New York, USA.

Topp R, Woolley S, Horuyak J, Khuder S, Kahaleh B. 2002. The effect of dynamic versus isometric resistance training on pain and functioning among adults with osteoarthritis of the knee. *Archives of Physical and Medical Rehabilitation*. 83:1187–95.

Topp R, Woolley S, Khuder S, Hornyak J, Bruss A. 2000, Predictors of Four Functional Tasks in Patients with Osteoarthritis of the Knee. *Orthopaedic Nursing*. 19(5): 49-59.

Torrance GW. 1976. Social preferences for health states: An empirical evaluation of three measurement techniques. *Socio-Econs Plan Sci* 10: 128-136.

Torrance GW. 1987. Utility approach to measuring health-related quality of life. *J Chronic Dis*. 40:593–600.

Towheed TE. 2006. Pennsaid Therapy for Osteoarthritis of the Knee: A Systematic Review and Metaanalysis of Randomized Controlled Trials. Journal of Rheumatology 33(3): 567 – 573.

Towheed T, Maxwell L, Judd M, Catton M, Hochberg MC, Wells GA. 2006. Acetaminophen for osteoarthritis. Cochrane Database of Systematic Reviews, Issue 1. Art. No.: CD004257. DOI: 10.1002/14651858.CD004257.pub2.

Towheed TE, Maxwell L, Anastassiades TP, Shea B, Houpt J, Robinson V, et al. 2005. Glucosamine therapy for treating osteoarthritis. *Cochrane Database of Systematic Reviews*, 2: CD002946.

TreeAge Pro 2011 TreeAge Software Inc, *Williamstown*, MA, USA.

Tu JV, Austin PC, Walld R, Roos L, Agras J, McDonald KM. 2001. Development and validation of the ontario acute myocardial infarction mortality prediction rules. *J. Am. Coll. Cardiol.* 37;992-997.

Urwin M, Symmons D, Allison T, Brammah T, Busby H, Roxby M, Simmons A, Gareth Williams G**.** 1998 Estimating the burden of musculoskeletal disorders in the community: the comparative prevalence of symptoms at different anatomical sites, and the relation to social deprivation. *Ann Rheum Dis.* 57:649–55.

van Baar M, Dekker J, Lemmens A, Oostendorp R, Bijlsma J. 1998a. Pain and disability in patients with osteoarthritis of the hip or knee: The relationship with articular, kinesiological, and psychological characteristics. *J. Rheumatol*. 25:125 – 33.

van Baar ME, Dekker J, Oostendorp RAB, Bijl D, Voorn TB, Lemmens JAM, Bijlsma JWJ. 1998b. The effectiveness of exercise therapy in patients with osteoarthritis of the hip or knee: a randomized clinical trial. *Journal of Rheumatology*. 25:2432–9.

Van Houwelingen JC, le Cessie. 1990. Predictive value of statistical models. *Statistics in medicine*. 8:1303-1325.

Vaughn TS and Berry KE. 2005. Using Monte Carlo techniques to demonstrate the meaning and implications of multicollinearity. Journal of Statistics Education, 13(1). http://www.amstat.org/publications/jse/v13n1/vaughan.html Accessed on 21-04-10.

Verhulst PF. 1838. "Notice sur la loi que la population poursuit dans son accroissement" (PDF). Correspondance mathématique et physique 10:113–121. http://books.google.com/?id=8GsEAAAAYAAJ&q=. Retrieved 09/08/2009.

Wang T, Belza B, Thompson EF, Whitney JD, Bennett K. 2007. Effects of aquatic exercise of flexibility, strength and aerobic fitness in adults with osteoarthritis of the hip or knee. *Journal of Advanced Nursing*. 57 (2), 141-152.

Ware JE, Kosinski M, Keller SD. 1996. A 12 item Short Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care*. 34:220 – 233.

Ware JE, Sherbourne CD. 1992. The MOS 36-items short-form health survey (SF-36) I. Conceptual framework and item selection. *Med Care*. 30:473 – 483.

Ware JE, Snow K, Kosinski M, Gandek B. 1993. SF-36 Health Survey. Manual and interpretation guide. Boston: *Health Institute,* New England Medical Centre.

Weinman J, Petrie KJ, Moss-Morris R, Horne R. 1996. The illness perception questionnaire: a new method for assessing the cognitive representation of illness. *Psychol Health.*11:431-445.

Weinstein M, Stason W. 1977. Foundations of cost effectiveness analysis for health and medical practices. *New Eng Jour of Med.* 296: 716-21.

Weinstein MC, O'Brien B, Hornberger J et al. 2003. Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices-Modeling Studies. Value Health 6:9–17.

Wetzels R, Weel CV, Grol R, Wensing M. 2008. Family practice nurses supporting self-management in older patients with mild osteoarthritis: a randomized trial. *BMC Family Practice*. 9:7.

Whitehurst DGT, Bryan S, Hay EM, Thomas E, Young J, Foster NE. 2011. Cost-effectiveness of acupuncture care as an adjunct to exercise-based physical therapy for osteoarthritis of the knee. *Phys Ther.* 91:630–641.

Wilder FV, Hall BJ, Barrett Jr JP et al. 2002. History of acute knee injury and osteoarthritis of the knee: a prospective epidemiological assessment. The Clearwater Osteoarthritis Study. *Osteoarthritis and Cartilage*. 10(8):611–6.

Wilkie R, Peat G, Croft P. 2002. Keele Assessment of Participation (KAP): A new participation instrument to measure the impact of osteoarthritis on the individual**.** *Rheumatology*, 41 (Suppl 2):22.

Wilkie R, Peat G, Thomas E, Croft P. 2007. Factors associated with participation restriction in community-dwelling adults aged 50 years and over. *Qual Life Res.* 16:1147-56.

Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. 1998. Prediction of coronary heart disease using risk factor categories. *Circulation.* 97(18): 1837-47.

World Health Organization. Preventing and managing the global epidemic of obesity. Report of the World Health Organization consultation on obesity. Geneva: World Health Organization, 1997.

World Health Organization: Global; Global Economic and Health Care Burden of Musculoskeletal Disease. 2001a. www.boneandjointdecade.org [Accessed on 17-08-09]

World Health Organization: World Health Report Archives 1995-2000. 2001. http://www.who.int/whr2001/2001/archives/1997/factse.htm. [Accessed on 17-08-09]

Yelin E, Callahan LF. 1995. The economic cost and social and psychological impact of musculoskeletal conditions. National Arthritis Data Work Groups. *Arthritis Rheum.* 38(10):1351-62.

Yelin EH. The economic and social burden of rheumatic disease. Chapter 5, pp. 5.1-54.In: Rheumatology. J Klippel and P Dieppe, eds. 2nd edition,1998

Yoshimura N, Campbell L, Hashimoto T, Kinoshita H, Okayasu T, Wilman C, Coggon D, Croft P, Cooper C. 1998. Acetabular dysplasia and hip osteoarthritis in Britan and Japan. *Bri. J. Rheum.* 37:1193 – 1197.

Your Total Health – Osteoarthritis: Signs and symptoms of osteoarthritis, Review on 10-01-2008 by Vikas Garg. http://yourtotalhealth.ivillage.com/osteoarthritis.html [Accessed on 27-10-09].

Yusuf E, Bijsterbosch J, Slagboom PE, Kroon HM, Rosendaal FR, et al. 2011. Association between Several Clinical and Radiological Determinants with Long-Term Clinical Progression and Good Prognosis of Lower Limb Osteoarthritis. PLoS ONE 6(10): e25426. doi:10.1371/journal.pone.0025426.

Zelman DC, Hoffman DL, Seifeldin R, Dukes EM. 2003. Development of a metric for a day of manageable pain control: derivation of pain severity cut-points for low back pain and osteoarthritis. *Pain*, 106:35 – 42.

Zhang J, Yu KF. 1998. What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. *JAMA.* 280:1690 –1691.

Zhang W, Daniel F McWilliams DF, Ingham SL, Sally A, Doherty SA, Muthuri S, Muir KR, Michael Doherty M. 2011. Nottingham knee osteoarthritis risk prediction models. *Ann Rheum Dis.* 70:1599 -1604.

Zhang W, Doherty M, Arden N, Bannwarth B, Bijlsma J, Gunther KP, Häuselmann HJ, Herrero-Beaumont G, Jordan K, Kaklamanis P, Leeb B, Lequesne M, Lohmander S, Mazieres B, Martin-Mola E, Pavelka K, Pendleton A, Punzi L, Swoboda B, Varatojo R, Verbruggen G, Zimmermann-Gorska I, Dougados M. 2005. EULAR evidence based recommendations for the management of hip osteoarthritis: report of a task force of the EULAR Standing Committee for International Clinical Studies Including Therapeutics (ESCISIT)**.** *Ann Rheum Dis.* 64: 669 - 681.

Zhang W, Doherty M, Leeb BF, Alekseeva L, Arden NK, Bijlsma JW, Dincer F, Dziedzic K, Häuselmann HJ, Herrero-Beaumont G, Kaklamanis P, Lohmander S, Maheu E, Martín-Mola E, Pavelka K, Punzi L, Reiter S, Sautner J, Smolen J, Verbruggen G, Zimmermann-Górska I. 2007a. EULAR evidence based recommendations for the management of hand osteoarthritis: Report of a Task Force of the EULAR Standing Committee for International Clinical Studies Including Therapeutics (ESCISIT) *Ann Rheum Dis.* 66: 377 - 388.

Zhang W, Jones A, Doherty M. 2004. Does paracetamol (acetaminophen) reduce the pain of osteoarthritis? A meta-analysis of randomized controlled trials. *Ann Rheum Disease.* 63(8): 901 – 907.

Zhang W, Moskowitz RW, Nuki G, Abramson S, Altman RD, Arden N, et al. 2007. OARSI recommendations for the management of hip and knee OA, part I: critical appraisal of existing treatment guidelines and systematic review of current research eveidence. Osteoarthritis Cartilage. 15: 981 – 1000.

Zhang W, Moskowitz RW, Nuki G, Abramson S, Altman RD, Arden N, et al. 2008. OARSI recommendations for the management of hip and knee OA, part II: OARSI evidence-based, expert consensus guidelines. Osteoarthritis Cartilage. 16: 137 – 62.

Zhang Y, Niu J, Kelly-Hayes M, Chaisson CE, Aliabadi P, Felson DT. 2002. Prevalence of symptomatic hand osteoarthritis and its impact on functional status among elderly: the Framingham study. *Am J Epidemiol.* 156:1021-7.

Zhang Y, Xu L, Nevitt MC, Aliabadi P, Yu W, Qin M, et al. 2001. Comparison of the prevalence of knee osteoarthritis between the elderly Chinese population in Beijing and whites in the United States: the Beijing Osteoarthritis Study. *Arthritis Rheum.* 44: 2065–71.

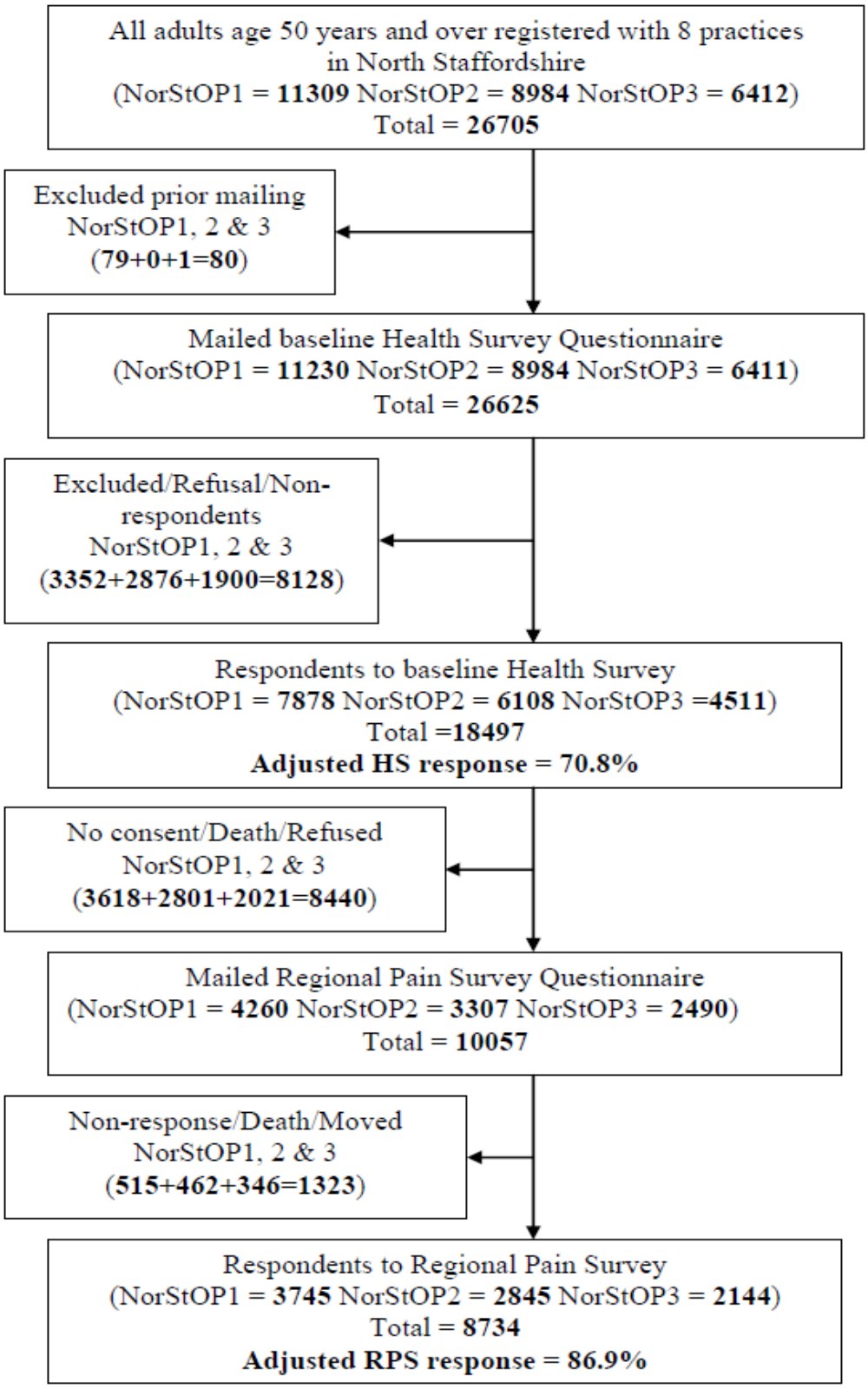
Zhao LP, Kolonel LN. 1992. Efficiency loss from categorizing quantitative exposures into qualitative exposures in case-control studies. *Am J Epidemiol.* 136: 464 – 474.

Zigmond AS, Snaith RP. 1983. The Hospital Anxiety and Depression Scale. *Acta Psychiatr Scand*. 6 7:361–70.

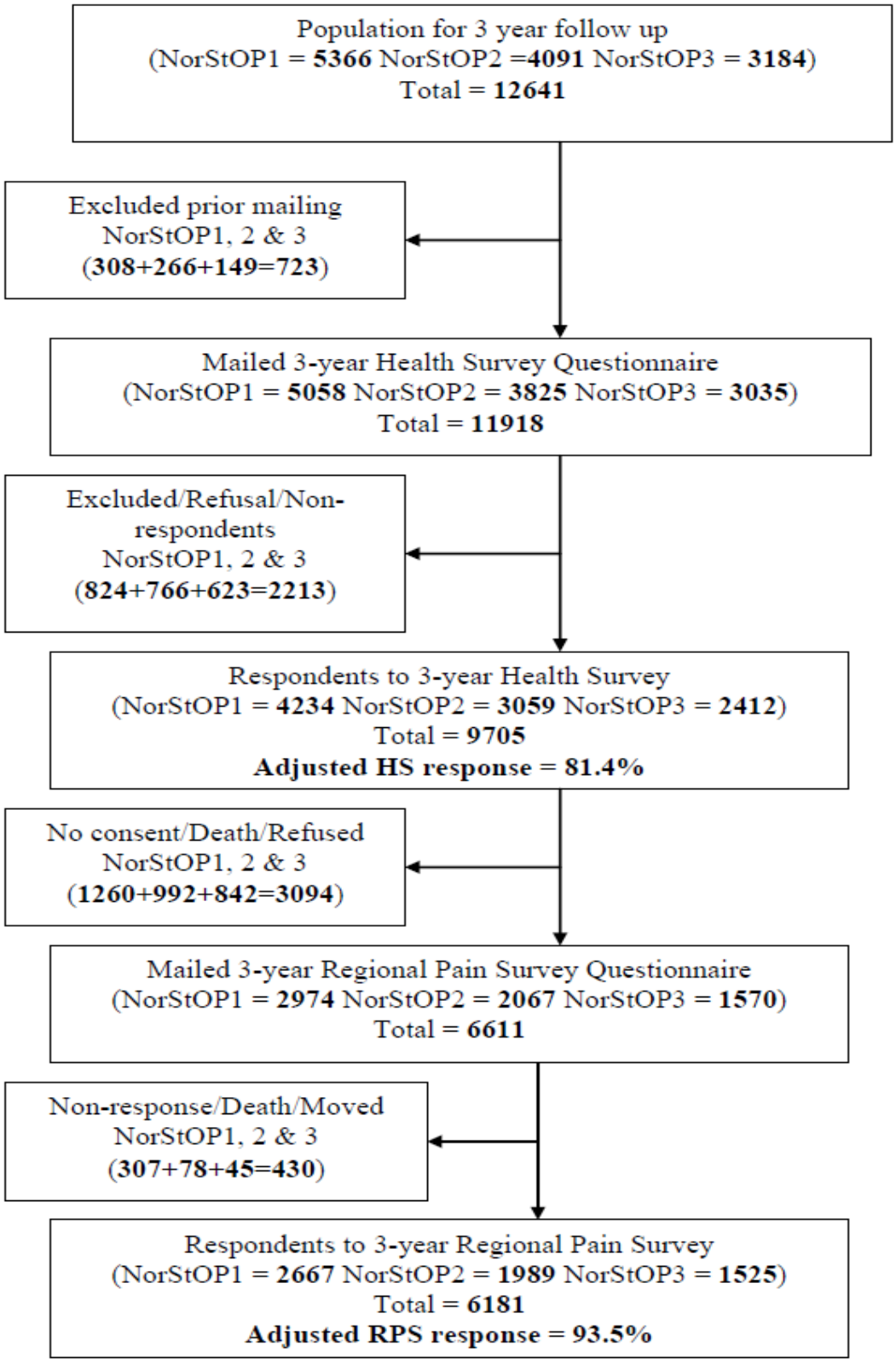
Zou G. 2004. A Modified Poisson Regression Approach to Prospective Studies with Binary Data. *American Jour Epid.*159(7): 702-706.

Appendices

Appendix 1a. Flowchart of recruitment into the NorStOP 1, 2 and 3 cohorts at baseline.



Appendix 1b. Flowchart of recruitment into the NorStOP 1, 2 and 3 cohorts at 3 year follow up.



Appendix 2. Description of baseline variables retained in the Poisson models among binary categories of severe pain and functional limitation at three years

	Severe pain at 3 years		Functional limitation at 3 years	
	Mild or No	Severe	Good	Poor
	N=1019(29%)	N=2544(71%)	N=1757(53%)	N=1535(47%)
Variables	n(%)	n(%)	n(%)	n(%)
†**Age Group**				
50 - 64 years	556(54.6)	1306(51.3)	1126(64.1)	641(41.8)
65 - 74 years	322(31.6)	852(33.5)	512(29.0)	557(36.3)
75 Plus years	141(13.8)	386(15.2)	119(6.9)	337(22.0)
†*Age [Median (IQR)]	63(56 - 70)	64(57 - 71)	62(56 - 67)	67(59 - 74)
Gender				
Male	452(44.4)	988(38.8)	766(43.6)	587(38.2)
Female	567(55.6)	1556(61.2)	991(56.4)	948(61.8)
Marital Status				
Married	788(78.2)	1810(71.8)	1394(80.0)	1023(67.4)
Unmarried	220(21.8)	711(28.2)	349(20.0)	494(32.6)
Employment Status				
Employed	314(31.5)	584(23.7)	680(39.0)	169(11.5)
Retired	538(54.0)	1352(54.9)	794(46.0)	925(62.8)
Unemployed	145(14.5)	525(21.3)	251(15.0)	378(25.7)
Social Class				
High	260(26.8)	454(18.8)	429(25.0)	248(17.2)
Low	709(73.2)	1961(81.2)	1265(75.0)	1191(82.8)
BMI				
Normal weight	383(39.4)	728(29.6)	653(38.0)	375(25.5)
Over weight	452(46.5)	1052(42.8)	780(46.0)	619(42.2)
Obese	137(14.1)	678(27.6)	275(16.0)	474(32.3)
Go onto full time education after school				
No	826(82.4)	2216(88.8)	1454(84.0)	1361(90.0)
Yes	177(17.7)	280(11.2)	276(16.0)	152(10.0)
Raised blood pressure				
No	716(70.3)	1557(61.2)	1258(71.6)	847(55.2)
Yes	303(29.7)	987(38.8)	499(28.4)	688(44.8)

† - Not significant at $p<0.05$ for severe pain but is significant for functional limitation
*Mann -Whitney U Test was used to determine p-value

Appendix 2. continued

	Severe pain at 3 years		Functional limitation at 3 years	
	Mild or No N=1019(29%)	Severe N=2544(71%)	Good N=1757(53%)	Poor N=1535(47%)
Variables	**n(%)**	**n(%)**	**N(%)**	**n(%)**
SF-36 Physical function score				
High (Good)	790(77.5)	1117(43.9)	1464(83.3)	311(20.3)
Low (Poor)	229(22.5)	1427(56.1)	293(16.7)	1224(79.7)
Front right foot (Man38)				
No	846(83.0)	1691(66.5)	1377(78.4)	972(63.3)
Yes	173(17.0)	853(33.5)	380(21.6)	563(36.7)
Hip pain in last year				
No	622(61.5)	1098(43.8)	1005(57.6)	592(39.0)
Yes	389(38.5)	1412(56.3)	739(42.4)	926(61.0)
WOMAC knee pain at Baseline				
Low	447(43.9)	680(26.7)	-	-
High	572(56.1)	1864(73.3)	-	-
Painkillers in last 4 wks				
All or most days	243(25.0)	1349(54.4)	-	-
Some days	204(20.9)	523(21.0)	-	-
Few or no days	527(54.1)	610(24.6)	-	-
Knee pain in last year				
No	386(38.1)	507(20.1)	-	-
Yes	627(61.9)	2011(79.9)	-	-
†**Access to advice or help with income**				
No	234(23.4)	619(24.9)	-	-
Yes	766(76.6)	1871(75.1)	-	-
AUSCAN Physical function				
Low (Better)	536(52.6)	737(29.0)	-	-
High(Poor)	483(47.4)	1807(71.0)	-	-
HADS Anxiety				
Low (Little distress)	577(56.6)	1024(40.3)	-	-
High (Most distress)	442(43.4)	1520(59.8)	-	-
Hand pain last year				
No	415(41.2)	682(27.1)	-	-
Yes	592(58.8)	1837(72.9)	-	-

All variables were significant at $p<0.05$ for both pain and functional limitation using Chi-square test

Appendix 2. continued

	Severe pain at 3 years		Functional limitation at 3 years	
	Mild or No N=1019(29 %)	Severe N=2544(71%)	Good N=1757(53%)	Poor N=1535(47%)
Variables	**n(%)**	**n(%)**	**n(%)**	**n(%)**
Natural remedies last 4 wks				
All of most days	341(37.7)	986(44.9)	-	-
Some days	50(5.5)	109(5.0)	-	-
Few or no days	514(56.8)	1102(50.2)	-	-
Go to a club, church or social event				
Most days in a week	89(8.9)	131(5.3)	-	-
Few days in a week	612(60.8)	1447(58.2)	-	-
No day in a week	305(30.3)	910(36.6)	-	-
When one goes to the doctor when ill				
Immediate or wait for few day	331(32.8)	916(36.4)	-	-
Wait several days	380(37.7)	811(32.2)	-	-
Put off as long as possible	298(29.5)	790(31.4)	-	-
AUSCAN pain				
Low	498(48.9)	731(28.7)	-	-
High	521(51.1)	1813(71.3)	-	-
Go out for a walk				
Most days in a week	-	-	752(43.6)	352(23.4)
Few days in a week	-	-	808(46.8)	656(43.7)
No day in a week	-	-	166(9.6)	495(32.9)
Back neck (man 43)				
No	-	-	1269(72.2)	924(60.2)
Yes	-	-	488(27.8)	611(39.8)
Back right shoulder(man 7)				
No	-	-	1369(77.9)	955(62.2)
Yes	-	-	388(22.1)	580(37.8)
Trouble falling asleep				
Not at all	-	-	700(40.6)	414(27.3)
On some nights	-	-	836(48.4)	741(48.9)
On most nights	-	-	190(11.0)	360(23.8)
HADS Depression				
Low	-	-	973(55.4)	386(25.1)
High	-	-	784(44.6)	1149(74.9)

Variables were significant at $p<0.05$ for pain and functional limitation using Chi-square test

Appendix 2. continued

	Severe pain at 3 years		Functional limitation at 3 years	
	Mild or No N=1019(29%)	Severe N=2544(71%)	Good N=1757(53%)	Poor N=1535(47%)
Variables	**n(%)**	**n(%)**	**n(%)**	**n(%)**
Go shopping				
Most day in a week	-	-	363(21.1)	227(15.0)
Few days in a week	-	-	1324(76.8)	1181(78.3)
No day in a week	-	-	37(2.1)	101(6.7)
WOMAC hip physical function at baseline				
Low	-	-	576(32.8)	282(18.4)
High	-	-	1181(67.2)	1253(81.6)
My health is very Unpredictable				
Agree and strongly agree	-	-	443(25.5)	790(52.3)
Disagree and strongly disagree	-	-	722(41.5)	301(19.9)
Neither agree or disagree	-	-	573(33.0)	420(27.8)
Compare to 12 months ago, have you reduced time or change how you have done any activity				
No, not at all	-	-	740(42.6)	150(9.9)
Yes, a lot	-	-	118(6.8)	551(36.5)
Yes, a little	-	-	879(50.6)	808(53.6)
Front right hip (man 46)				
No	-	-	1497(85.2)	1096(71.4)
Yes	-	-	260(14.8)	439(28.6)
Doctors can do a lot to help people with joint problems				
Agree and strongly agree	-	-	1110(63.5)	1088(71.7)
Disagree and strongly disagree	-	-	173(9.9)	143(9.4)
Neither agree or disagree	-	-	146(26.6)	287(18.9)
Knee problem last year				
No	-	-	619(35.6)	245(16.2)
Yes	-	-	1122(64.4)	1271(83.8)
Look after others				
Most day in a week	-	-	580(33.8)	369(24.6)
Few days in a week	-	-	522(30.4)	336(22.4)
No day in a week	-	-	616(35.8)	793(53.0)

All variables were significant at $p<0.05$ for both pain and functional limitation using Chi-square test

Appendix 2. continued

	Severe pain at 3 years		Functional limitation at 3 years	
	Mild or No N=1019(29%)	Severe N=2544(71%)	Good N=1757(53%)	Poor N=1535(47%)
Variables	**N(%)**	**n(%)**	**n(%)**	**n(%)**
Front left elbow (man 29)				
No	-	-	1657(94.3)	1329(86.6)
Yes	-	-	100(5.7)	206(13.4)
Cost of living				
Quite comfortable	-	-	332(19.1)	150(10.0)
Strain	-	-	34(2.0)	103(6.8)
Have to be careful	-	-	595(34.1)	758(50.3)
Able to manage	-	-	782(44.8)	496(32.9)
SF-12 Physical component score at baseline				
High (Good)	-	-	1392(79.2)	411(26.8)
Low (Poor)	-	-	365(20.8)	1124(73.2)
I have power to influence what happens in my life				
Agree and strongly agree	-	-	1118(64.0)	869(57.3)
Disagree and strongly disagree	-	-	192(11.0)	249(16.4)
Neither agree or disagree	-	-	435(25.0)	398(26.3)
In past 4 weeks, have you reduced time or change how you have done any activity because of health				
Most days in a week	-	-	100(5.8)	500(33.0)
Few days in a week	-	-	685(39.4)	771(50.9)
No day in a week	-	-	951(54.8)	243(16.1)
Trouble staying asleep				
Not at all	-	-	544(32.0)	289(19.3)
On some nights	-	-	828(48.8)	719(48.1)
On most nights	-	-	326(19.2)	488(32.6)

All variables were significant at $p<0.05$ for functional limitation using Chi-square test

Appendix 3. Final logistic regression model for severe pain at three years

Variables	N	OR(95% CI)	Adjusted PAR(95% CI)	Unadjusted PAR(95% CI)	Unadjusted NNT(95% CI)
Knee pain in last year					
No	893	1			
Yes	2638	2.19(1.57, 3.05)	12.1(7.6, 16.4)	18.6(5.1, 32.2)	5.1(4.3, 6.3)
AUSCAN Physical function					
Low	1273	1			
High	2290	1.83(1.41, 2.37)	8.4(5.6, 11.2)	13.2(5.5, 20.9)	4.8(4.1, 5.6)
WOMAC knee pain at baseline					
Low	1127	1			
High	2436	1.79(1.35, 2.38)	8.8(5.6, 12.0)	12.7(2.2, 23.2)	6.2(5.1, 7.8)
Trouble falling asleep					
Not at all	1188	1			
On most nights	1718	1.75(1.09, 2.82)	1.5(0.6, 2.3)	3.3(-6.5, 13.0)	4.2(3.5, 4.9)
Front left shin (man 41)					
No	2945	1			
Yes	618	1.73(1.09, 2.76)	0.8(-0.1, 1.7)	2.6(1.0, 4.2)	6.7(5.4, 8.6)
BMI					
Normal weight	1111	1			
Obesity	815	1.72(1.23, 2.40)	2.2(1.0, 3.4)	3.8(-8.4, 16.1)	5.7(4.7, 7.2)
Hand pain in last year					
No	1097	1			
Yes	2429	1.63(1.25, 2.11)	7.8(4.7, 10.9)	10.5(-1.1, 22.2)	7.4(6.0, 9.9)
SF-36 Physical function score					
High (Good)	1907	1			
Low (Poor)	1656	1.63(1.19, 2.22)	4.1(2.3, 6.0)	8.1(4.5, 11. 7)	3.6(3.3, 4.0)
Raised blood pressure					
No	2273	1			
Yes	1290	1.50(1.14, 1.97)	1.7(0.2, 3.2)	4.1(0.6, 7.6)	12.5(9.1, 20.0)

Appendix 3. continued

Variables	N	OR(95% CI)	Adjusted PAR(95% CI)	Unadjusted PAR(95% CI)	Unadjusted NNT(95% CI)
Front right knee (man 36)					
No	1924	1			
Yes	1639	1.49(1.09, 2.06)	2.1(-0.1, 4.2)	5.9(1.7, 10.2)	5.9(5.0, 7.1)
Back right hip (man 45)					
No	2801	1			
Yes	762	1.47(1.01, 2.14)	0.9(-0.1, 1.9)	2.4(0.4, 4.3)	7.5(6.0, 9.9)
Foot pain in last year					
No	1520	1			
Yes	1999	1.45(1.11, 1.90)	3.9(1.5, 6.3)	7.0(0.5, 13.5)	6.4(5.4, 8.0)
Go onto full time education after school					
Yes	457	1			
No	3042	1.44(1.03, 2.00)	6.2(0.6, 11.5)	7.2(-40.1, 54.6)	8.6(6.1, 14.6)
FPDI pain					
No	1089	1			
Yes	2474	1.42(1.06, 1.90)	4.8(1.3, 8.2)	7.2(-6.5, 20.8)	12.9(9.1, 22.5)
Hip pain in last year					
No	1720	1			
Yes	1801	1.38(1.05, 1.80)	3.6(1.4, 5.7)	5.3(-0.1, 10.7)	6.9(5.7, 8.6)
Access to advice or help with income					
No	853	1			
Yes	2637	1.37(1.01, 1.88)	0.9(-0.3, 2.1)	1.9(-0.5, 4.4)	61.9(19.7, -54.2)
HADS Anxiety					
Low	1601	1			
High	1962	1.31(1.02, 1.68)	1.5(-0.8, 3.8)	4.8(-1.5, 11.1)	7.4(6.1, 9.5)
When one goes to the doctor when ill					
Immediate or wait for few day	1247	1			
Wait several days	1191	0.77(0.59, 0.99)	-	-	-

Appendix 3. continued

Variables	N	OR(95% CI)	Adjusted PAR(95% CI)	Unadjusted PAR(95% CI)	Unadjusted NNT(95% CI)
The thought of pain makes me afraid					
Agree and strongly agree	978	1			
Neither agree or disagree	1011	0.74(0.57, 0.97)	-	-	-
Natural remedies last 4 wks					
All of most days	1327	1			
Few or no days	1616	0.72(0.55, 0.92)	-	-	-
Go to a club, church or social event					
Most days in a week	220	1			
No day in a week	1215	0.68(0.51, 0.90)	-	-	-
Front right shin (man 37)					
No	2832	1			
Yes	731	0.65(0.43, 0.99)	-	-	-
Painkillers in last 4 wks					
All or most days	1592	1			
Some day	727	0.60(0.42, 0.85)	-	-	-
Few or no day	1137	0.43(0.32, 0.59)	-	-	-
Joint problems always gets worse overtime					
Agree and strongly agree	2865	1			
Disagree or strongly disagree	125	0.54(0.31, 0.97)	-	-	-

Pearson goodness of fit chi-square test = 2582.49, p-value = 0.686
C-Statistic or Area under ROC = 0.793(0.775 to 0.811)
Total number of subjects used to derive the model = 1643
N – Number of subjects

IRR(95% CI) – Incident rate ratio (95% Confidence Interval)

PAR – Population Attributable Risk

NNT – Number Needed to Treat

Man – Body manikin: a tool made up of 50 items that covers the whole body used to measure bodily pain.

Appendix 4. Final logistic regression model for functional limitation at three years

Variables	N	OR(95% CI)	Adjusted PAR(95% CI)	Unadjusted PAR(95% CI)	Unadjusted NNT(95% CI)
Physical function (SF-36) score at baseline					
High (Good)	1907	1			
Low (Poor)	1656	5.08(3.61, 7.14)	29.4(23.4, 35.0)	48.8(45.1, 52.4)	1.7(1.7, 1.8)
Cost of living					
Quite comfortable	521	1			
Strain	154	3.35(1.37, 8.21)	1.0(-2.7, 4.6)	4.2(-19.1, 27.5)	2.6(2.2, 3.4)
Physical component (SF-12) Score					
High (Good)	1952	1			
Low (|Poor)	1611	2.40(1.70, 3.38)	13.3(8.7, 17.8)	27.8(24.4, 31.3)	2.1(1.9, 2.2)
Require assistance to go places					
No	2719	1			
Yes	783	2.30(1.50, 3.54)	3.3(1.2, 5.2)	12.2(10.2, 14.3)	2.1(2.0, 2.2)
Go onto full time education after school					
Yes	457	1			
No	3042	2.21(1.36, 3.58)	11.9(3.1, 20.0)	5.4(-32.8, 43.5)	8.7(6.2, 14.7)
Go out for a walk					
Most day in a week	1184	1			
No day in a week	720	2.12(1.45, 3.11)	3.3(1.0, 5.5)	10.9(3.6, 18.1)	2.6(2.3, 2.9)
Front left elbow (Man 29)					
No	3226	1			
Yes	219	2.10(1.35, 3.26)	1.2(0.1, 2.3)	4.0(2.9, 5.2)	5.0(3.9, 6.9)
Current employment status					
Employed	898	1			
Retired	1890	1.85(1.37, 2.50)	12.4(7.6, 17.0)	24.5(17.5, 31.5)	3.3(3.0, 3.7)
Access to car when personally need it					
Yes	2963	1			
No	547	1.78(1.09, 2.89)	2.8(1.4, 4.2)	16.3(14.7, 18.0)	6.1(4.8, 8.3)

Appendix 4. continued

Variables	N	OR(95% CI)	Adjusted PAR(95% CI)	Unadjusted PAR(95% CI)	Unadjusted NNT(95% CI)
Walks of two miles or more					
Most day in a week	294	1			
No day in a week	2030	1.76(1.27, 2.45)	8.4(-0.9, 16.8)	23.6(8.1, 39.1)	2.6(2.3, 3.1)
Knee pain last year					
No	893	1			
Yes	2638	1.74(1.23, 2.47)	12.8(6.6, 18.6)	25.2(15.7, 34.8)	4.6(4.0, 5.5)
Trouble staying asleep					
Not at all	889	1			
On some nights	1659	1.69(1.24, 2.30)	6.1(1.9, 10.1)	15.4(2.9, 28.0)	9.2(6.8, 14.4)
WOMAC hip function at baseline					
Low	904	1			
High	2659	1.58(1.09, 2.30)	4.8(-1.3, 10.6)	20.2(8.2, 32.3)	6.3(5.1, 8.1)
Back right shoulder (man7)					
No	2531	1			
Yes	1032	1.56(1.08, 2.26)	3.9(1.2, 6.5)	7.7(5.2, 10.2)	5.1(4.5, 6.7)
Hip pain last year					
No	1720	1			
Yes	1801	1.50(1.06, 2.14)	2.3(-1.7, 6.2)	12.6(7.7, 17.4)	5.9(5.0, 7.3)
Raised blood pressure					
No	2273	1			
Yes	1290	1.47(1.09, 1.99)	2.5(0.1, 5.0)	8.4(5.2, 11.5)	6.2(5.1, 7.9)
BMI					
Normal weight	1111	1			
Obesity	815	1.46(1.03, 2.09)	2.3(0.1, 4.4)	5.9(-3.7, 15.6)	4.1(3.5, 5.0)
WOMAC knee function at baseline					
Low	904	1			
High	2659	1.45(1.04, 2.02)	8.0(3.1, 12.6)	16.1(10.4, 21.8)	3.7(3.3, 4.1)
Front right foot (man 38)					
No	2537	1			
Yes	1026	1.41(1.02, 1.96)	1.6(-0.6, 3.8)	6.0(3.4, 8.5)	6.0(5.0, 7.7)

Appendix 4. continued

Variables	N	OR(95% CI)	Adjusted PAR(95% CI)	Unadjusted PAR(95% CI)	Unadjusted NNT(95% CI)
Doctors can do a lot to help people with joint problems					
Agree and strongly agree	2391	1			
Neither agree or disagree	809	0.70(0.50, 0.99)	-	-	-
Front right hip (man 46)					
No	2812	1			
Yes	751	0.66(0.45, 0.98)	-	-	-
My health is very unpredictable					
Agree and strongly agree	1355	1			
Disagree and strongly disagree	1092	0.61(0.44, 0.85)	-	-	-
In past 4 weeks, have you reduced time or change how you have done any activity because of health					
Most day in a week	652	1			
No day in a week	1276	0.59(0.41, 0.85)	-	-	-
Trouble falling asleep					
Not at all	1188	1			
On some nights	1718	0.58(0.42, 0.79)	-	-	-
Back neck (man 43)					
No	2380	1			
Yes	1183	0.57(0.39, 0.83)	-	-	-
Go shopping					
Most day in a week	644	1			
No day in a week	152	0.43(0.20, 0.94)	-	-	-

Pearson goodness of fit chi-square = 2961.69, p-value = 0.058
C-Statistic or Area under ROC = 0.885(0.872 to 0.897)
Total number of subjects used to derive the model = 1602
N – Number of subjects

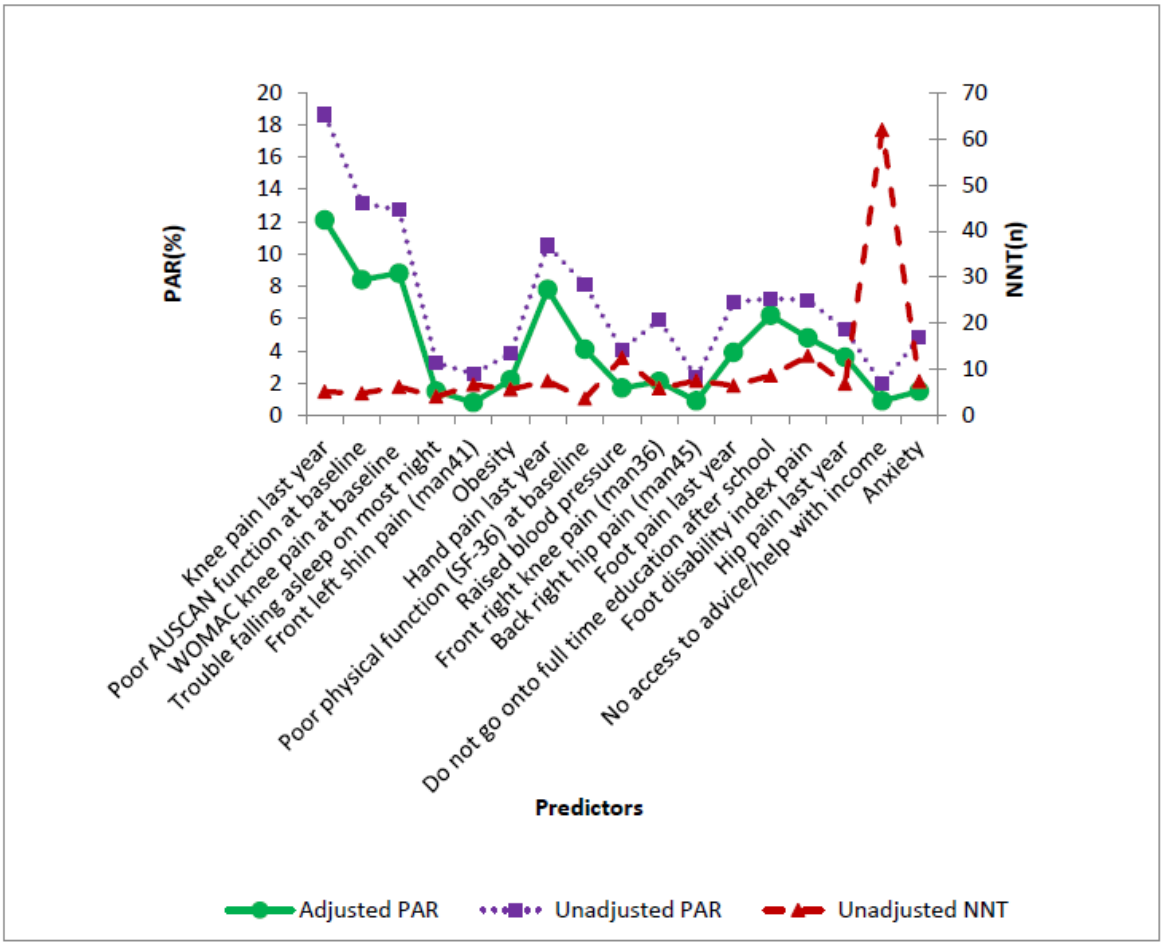
IRR(95% CI) – Incident rate ratio (95% Confidence Interval)

PAR – Population Attributable Risk

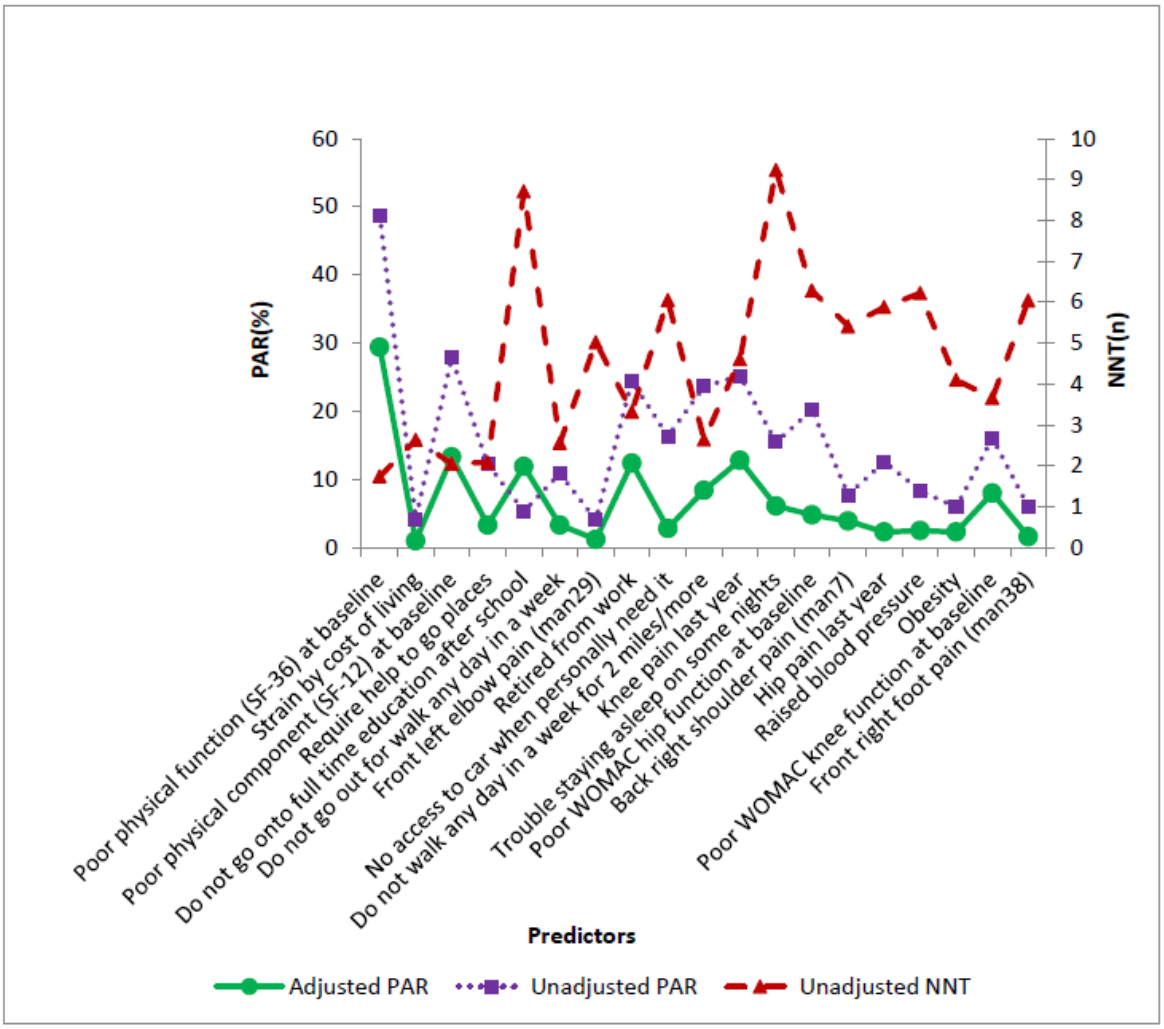
NNT – Number Needed to Treat

Man – Body manikin: a tool made up of 50 items that covers the whole body used to measure bodily pain.

Appendix 5a. Adjusted PAR, unadjusted PAR and unadjusted NNT for predictors associated with increase severe pain at three years in the final logistic regression model



Appendix 5b. Figure4.5a. Adjusted PAR, unadjusted PAR and unadjusted NNT for predictors associated with poor functional limitation at three years in the final logistic regression model



Appendix 6a. Comparison of the predictors of severe pain at three years selected in the MI Poisson regression models but not in the unimputed Poisson regression model and vice versa

Variables selected in MI model but not in unimputed model	Direction of risk
Foot disability index pain	Increased risk
Trouble falling asleep on most nights	Increased risk
Back right thigh pain	Increased risk
Trouble falling asleep on some nights	Increased risk
Wake up feeling tired on some nights	Increased risk
Spine pain (man2)	Increased risk
Put off as long as possible when ill before go to GP	Increased risk
Back right foot pain (man21)	Decrease risk
Neither dis(agree) if pain last for a week/more have serious disease	Decrease risk
Disagree if pain last for a week/more have serious disease	Decrease risk
No day in a week do home maintenance activities	Decrease risk
Few days in a week do home maintenance activities	Decrease risk
Foot disability index function	Decrease risk
Some days on natural remedies last 4 weeks	Decrease risk
Disagree joint problem always gets worse over time	Decrease risk
Variables selected in unimputed model but not in MI model	
Anxiety	Increased risk
No access to advice/help with income	Increased risk
Wait several days when ill before go to GP	Decrease risk
Do not go to club/church/social	Decrease risk

Appendix 6b. Comparison of the predictors of poor function at three years selected in the MI Poisson regression models but not in the unimputed Poisson regression model and vice versa

Variables selected in MI model but not in unimputed model	Direction of risk
No day in a week go out to work	Increased risk
Few days in a week go out to work	Increased risk
75 plus years	Increased risk
Do not have access to car when personally need it	Increased risk
65 - 74 years	Increased risk
No day in a week do heavy house work	Increased risk
Currently smoking	Increased risk
Front left shoulder pain (man28)	Increased risk
Foot problem last year	Increased risk
Few days in a week go in a car as a passenger	Increased risk
No day in a week walk for at least a quarter of a mile	Increased risk
Few days in a week cook and clean	Decrease risk
Female	Decrease risk
Spine pain (man2)	Decrease risk
Poor WOMAC hip stiffness at baseline	Decrease risk
No/few days on painkillers last 4 weeks	Decrease risk
Variables selected in unimputed model but not in MI model	
Retired from work	Increased risk
A lot time reduced time/change activities 1yr ago	Increased risk
Unemployed	Increased risk
Depression	Increased risk
Back right shoulder pain (man7)	Increased risk
Do not go out for walk any day in a week	Increased risk
Front left elbow pain (man29)	Increased risk
Strain by cost of living	Increased risk
Raised blood pressure	Increased risk
Front right foot pain (man38)	Increased risk
Front right hip pain (man46)	Decrease risk
Neither dis(agree) can influence what happens in life	Decrease risk
Look after others few days in a week	Decrease risk
Go shopping few days in a week	Decrease risk
Trouble falling asleep on most nights	Decrease risk
Back neck pain (man43)	Decrease risk

Appendix 6c. Comparison of the predictors of severe pain at three years selected in the MI logistic regression models but not in the unimputed logistic regression model and vice versa.

Variables selected in MI model but not in unimputed model	Direction of risk
Seldom/hardly ever visit GP for oneself	Increased risk
Occasionally visit GP for oneself	Increased risk
Spine pain (man2)	Increased risk
Trouble falling asleep on some nights	Increased risk
Few days in a week do home maintenance activities	Decrease risk
Neither dis(agree) if pain last for a week/more have serious disease	Decrease risk
Disagree if pain last for a week/more have serious disease	Decrease risk
Foot disability index function	Decrease risk
Past 4 weeks few days in a week reduced time/change activities	Decrease risk
Past 4 weeks no day in a week reduced time/change activities	Decrease risk
No/few days on painkillers last 4 weeks	Decrease risk
Disagree osteoarthritis is a serious condition	Decrease risk
Variables selected in unimputed model but not in MI model	
Front left shin pain (man41)	Increased risk
Back right hip pain (man45)	Increased risk
No access to advice/help with income	Increased risk
Anxiety	Increased risk
Wait several days when ill before go to GP	Decrease risk
Neither dis(agree) thought of pain makes me afraid	Decrease risk
No/few days on natural remedies last 4 weeks	Decrease risk
Do not go to club/church/social	Decrease risk
Front right shin pain (man37)	Decrease risk
Disagree joint problem always gets worse over time	Decrease risk
No/few days on painkillers last 4 weeks	Decrease risk

Appendix 6d. Comparison of the predictors of poor function at three years selected in the MI logistic regression models but not in the unimputed logistic regression model and vice versa

Variables selected in MI model but not in unimputed model	Direction of risk
Knee problems last year	Increased risk
Few days in a week go out to work	Increased risk
75 plus years	Increased risk
No day in a week go out to work	Increased risk
Back right hand pain (man10)	Increased risk
65 - 74 years	Increased risk
Front left shoulder pain (man28)	Increased risk
Front left hand pain (man31)	Increased risk
Back right foot pain (man21)	Increased risk
Heart problems	Increased risk
Few days in a week go in a car as a passenger	Increased risk
No day in a week do heavy house work	Increased risk
Medium/high isolation	Decrease risk
Put off as long as possible when ill before go to GP	Decrease risk
No/few days on painkillers last 4 weeks	Decrease risk
Back left upper torso pain (man11)	Decrease risk
Front right hand pain (man27)	Decrease risk
Poor WOMAC hip stiffness at baseline	Decrease risk
Back left hand pain (man6)	Decrease risk
Variables selected in unimputed model but not in MI model	
Strain by cost of living	Increased risk
Front left elbow pain (man29)	Increased risk
Retired from work	Increased risk
Knee pain last year	Increased risk
Back right shoulder pain (man7)	Increased risk
Hip pain last year	Increased risk
Raised blood pressure	Increased risk
Front right foot pain (man38)	Increased risk
Neither dis(agree) GP can help with joint problem	Decrease risk
Front right hip pain (man46)	Decrease risk
Back neck pain (man43)	Decrease risk
Do not go shopping in a week	Decrease risk

Appendix 7a. Search terms for primary care for OA in Cochrane database - 19 Aug 2010

ID	Search	Hits
#1	MeSH descriptor **Osteoarthritis** explode all trees	2728
#2	(osteoarthriti*):ti,ab,kw	4059
#3	(osteoarthro*)	399
#4	(gonarthriti*):ti,ab,kw	24
#5	(gonarthro*):ti,ab,kw	190
#6	(coxarthriti*):ti,ab,kw	4
#7	(coxarthro*):ti,ab,kw	94
#8	(arthros*):ti,ab,kw	1819
#9	(arthrot*):ti,ab,kw	80
#10	((knee* or hip* or hand* or foot* or joint*) near/3 (pain* or ach* or discomfort* or stiff*))	3060
#11	(#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10)	7737
#12	MeSH descriptor **Family Practice** explode all trees	2201
#13	MeSH descriptor **Physicians, Family** explode all trees	465
#14	(GP):ti,ab,kw	1329
#15	(general near/2 practi*):ti,ab,kw	4888
#16	(general near/2 physician*):ti,ab,kw	72
#17	(family near/2 physician*):ti,ab,kw	858
#18	(family near/2 practi*):ti,ab,kw	2631
#19	(family near/2 doctor*):ti,ab,kw	124
#20	MeSH descriptor **Primary Health Care** explode all trees	2921
#21	MeSH descriptor **Community Health Services** explode all trees	19440
#22	(primary near/2 care*):ti,ab,kw	6538
#23	(primary near healthcare*):ti,ab,kw	140
#24	(community near/2 (service* or care*)):ti,ab,kw	1795
#25	(community near healthcare):ti,ab,kw	36
#26	MeSH descriptor **Ambulatory Care** explode all trees	3581

Appendix 7a. continued

ID	Search	Hits
#27	(ambulatory near/2 care*):ti,ab,kw	3853
#28	(ambulatory near healthcare):ti,ab,kw	3
#29	(#12 OR #13 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR #21 OR #22 OR #23 OR #24 OR #25 OR #26 OR #27 OR #28)	33458
#30	MeSH descriptor **Anti-Inflammatory Agents, Non-Steroidal** explode all trees	12804
#31	MeSH descriptor **Antirheumatic Agents** explode all trees	22118
#32	(NSAID* OR anti-inflammatory OR antirheumatic):ti,ab,kw	13745
#33	(drug* OR agent*):ti,ab,kw	292032
#34	(exercise* OR physiotherap* OR "physical activity" OR " physical therapy"):ti,ab,kw	33749
#35	(**#30** OR **#31** OR **#32** OR **#33** OR **#34**)	317054
#36	(#29 OR #35)	338883
#37	(#11 AND #36)	4649
#38	(#11 AND #36), from 1990 to 2010	3690
#39	(#11 AND #36), from 2000 to 2010	2505

Appendix 7b. Search terms for primary care for OA in Medline database

No.	Database	Search term	Hits
1	MEDLINE	exp OSTEOARTHRITIS/	35571
2	MEDLINE	osteoarthriti*.ti,ab	28302
3	MEDLINE	osteoarthro*.ti,ab	4555
4	MEDLINE	gonarthriti*.ti,ab	130
5	MEDLINE	gonarthro*.ti,ab	824
6	MEDLINE	coxarthriti*.ti,ab	81
7	MEDLINE	coxarthro*.ti,ab	1318
8	MEDLINE	arthros*.ti,ab	19846
9	MEDLINE	arthrot*.ti,ab	1913
10	MEDLINE	(((knee* OR hip* OR hand* OR foot* OR joint*) adj3 (pain* OR ach* OR discomfort* OR stiff*))).ti,ab	19460
11	MEDLINE	1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10	80423
12	MEDLINE	exp FAMILY PRACTICE/	59688
13	MEDLINE	exp PHYSICIANS, FAMILY/	14304
14	MEDLINE	GP.ti,ab	22007
15	MEDLINE	((general adj2 practi*)).ti,ab	58187
16	MEDLINE	((family adj2 physician*)).ti,ab	10474
17	MEDLINE	((family adj2 practi*)).ti,ab	10465
18	MEDLINE	((family adj2 doctor*)).ti,ab	3405
19	MEDLINE	((general adj2 physician*)).ti,ab	2134
20	MEDLINE	exp PRIMARY HEALTH CARE/	62296
21	MEDLINE	exp COMMUNITY HEALTH SERVICES/	425365
22	MEDLINE	((primary adj2 care*)).ti,ab	67141
23	MEDLINE	((primary ADJ healthcare*)).ti,ab	1310
24	MEDLINE	((community adj2 (service* OR care*))).ti,ab	12332
25	MEDLINE	((community ADJ healthcare*)).ti,ab	169
26	MEDLINE	exp AMBULATORY CARE/	41284
27	MEDLINE	((ambulatory adj2 care*)).ti,ab	7502

Appendix 7b. continued

No.	Database	Search term	Hits
28	MEDLINE	((ambulatory ADJ healthcare)).ti,ab	35
29	MEDLINE	12 OR 13 OR 14 OR 15 OR 16 OR 17 OR 18 OR 19 OR 20 OR 21 OR 22 OR 23 OR 24 OR 25 OR 26 OR 27 OR 28	640874
30	MEDLINE	advi*.ti,ab	62152
31	MEDLINE	exp DIET THERAPY/	35524
32	MEDLINE	diet.ti,ab	173115
33	MEDLINE	exp ANTIRHEUMATIC AGENTS/	291408
34	MEDLINE	exp ANTI-INFLAMMATORY AGENTS/	356854
35	MEDLINE	(medicat* OR drug* OR agent*).ti,ab	1444648
36	MEDLINE	NSAID*.ti,ab	14533
37	MEDLINE	(anti-inflammator* OR antiinflammator*).ti,ab	75821
38	MEDLINE	(anti-rheumatic* OR antirheumatic*).ti,ab	5035
39	MEDLINE	exp EXERCISE THERAPY/	22433
40	MEDLINE	(exercise* OR physiotherap*).ti,ab	169438
41	MEDLINE	(physical ADJ therap*).ti,ab	10281
42	MEDLINE	(physical ADJ activit*).ti,ab	38640
43	MEDLINE	30 OR 31 OR 32 OR 33 OR 34 OR 35 OR 36 OR 37 OR 38 OR 39 OR 40 OR 41 OR 42	2246045
44	MEDLINE	11 AND 29 AND 43	855
45	MEDLINE	44 [Limit to: Publication Year 2000-Current]	573
46	MEDLINE	45 [Limit to: Publication Year 2005-Current and Publication Year 2000-Current]	573
47	MEDLINE	45 [Limit to: Publication Year 2005-Current and (Age Groups Middle Aged 45 plus years) and Publication Year 2000-Current]	340
48	MEDLINE	45 [Limit to: Publication Year 2003-Current and (Age Groups Middle Aged 45 plus years) and Publication Year 2000-Current]	340
49	MEDLINE	45 [Limit to: Publication Year 2002-Current and (Age Groups Middle Aged 45 plus years) and Publication Year 2000-Current]	340
50	MEDLINE	45 [Limit to: Publication Year 2001-Current and (Age Groups Middle Aged 45 plus years) and Publication Year 2000-Current]	340

Appendix 8a. Summary of mean and standard deviation scores of outcome for the advice and information studies

Author	Outcome	Treatment (n)	Treatment mean	Treatment sd	Control (n)	Control mean	Control sd
Keefe, 1990	Pain (AIMS)	36	-5.91	1.95	31	-5.64	1.79
Keefe, 1990	Function (AIMS)	36	-2.63	1.50	31	-1.96	1.43
Heuts, 2005	Pain (VAS)	132	3.70	2.60	141	4.20	2.70
Heuts, 2005	Function (SF-36)	132	-61.50	21.30	141	-55.40	22.80
Wetzels, 2008	Pain (AIMS)	51	11.19	3.95	53	11.48	3.64
Wetzels, 2008	Function (AIMS)	51	14.56	4.52	53	14.40	4.74
Ravaud, 2009	Pain (VAS)	146	-1.65	2.32	181	-1.18	2.58
Ravaud, 2009	Function (WOMAC)	146	-5.74	10.66	181	-4.03	11.35

n – Number of patients; Sd – Standard deviation; Negative mean estimate – Change score or change in the direction of reduce pain or function.

Appendix 8b. Summary of mean and standard deviation scores of outcome for the simple analgesia studies

Author	Outcome	Treatment (n)	Treatment mean	Treatment sd	Control (n)	Control mean	Control sd
Case, 2003	Pain (WOMAC)†	29	-186.90	121.50	28	-183.40	122.90
Case, 2003	Function (WOMAC)†	29	-615.20	360.20	28	-611.50	365.40
Pincus, 2004	Pain (VAS)	171	50.10	27.10	172	53.50	27.00

n – Number of patients; Sd – Standard deviation; Negative mean estimate – Change score or change in the direction of reduce pain or function. † – Used 100mm Visual analogue scale.

Appendix 8c. Summary of mean and standard deviation scores of outcome for the topical NSAIDs studies

Author	Outcome	Treatment (n)	Treatment mean	Treatment sd	Control (n)	Control mean	Control sd
Grace, 1999	Pain (WOMAC)	34	28.19	18.31	34	35.42	19.86
Grace, 1999	Function (WOMAC)	34	34.41	18.14	34	37.44	21.16
Bookman, 2004	Pain (WOMAC)	84	5.20	4.60	79	6.90	4.50
Bookman, 2004	Function (WOMAC)	84	17.90	15.60	79	23.70	15.90
Roth, 2004	Pain (WOMAC)	163	7.10	4.70	159	8.60	4.90
Roth, 2004	Function (WOMAC)	162	26.60	15.60	159	31.20	15.80
Niethard, 2005	Pain (WOMAC)	117	-22.00	21.00	120	-14.00	23.00
Niethard, 2005	Function (WOMAC)	117	-23.00	21.00	120	-16.00	22.00

n – Number of patients; Sd – Standard deviation; Negative mean estimate – Change score or change in the direction of reduce pain or function.

Appendix 8d. Summary of mean and standard deviation scores of outcome for exercise studies

Author	Outcome	Treatment (n)	Treatment mean	Treatment sd	Control (n)	Control mean	Control sd
Bautch, 1997	Pain (VAS)	15	-2.19	0.43	15	-2.08	0.54
Bautch, 1997	Function (AIMS)	15	-23.37	2.48	15	-17.88	1.85
Ettinger, 1997	Pain (FAST)	144	-0.32	0.60	75	0.00	0.61
Ettinger, 1997	Function (FAST)	144	-0.18	0.48	75	0.00	0.48
van Baar, 1998	Pain (VAS)	54	-27.40	28.70	59	-11.70	28.50
van Baar, 1998	Function (IRGL)	54	-1.30	5.70	59	-0.50	5.60
Maurer, 1999	Pain (WOMAC)†	49	-43.54	80.30	49	-28.49	80.30
Maurer, 1999	Function (WOMAC)†	49	-106.90	390.10	49	-88.30	390.10
O'Reilly, 1999	Pain (WOMAC)	108	-1.45	3.50	72	-0.42	2.80
O'Reilly, 1999	Function (WOMAC)	108	-3.55	12.50	72	-0.01	11.50
Peloquin, 1999	Pain (AIMS)	59	-1.44	2.00	65	-0.59	2.20
Peloquin, 1999	Function (AIMS)	59	-1.50	2.40	65	-0.54	2.60
Deyle, 2000	Pain (WOMAC)†	33	-129.63	91.00	36	-33.83	111.50
Deyle, 2000	Function (WOMAC)†	33	-402.51	339.56	36	-98.17	393.90
Hopman-Rock, 2000	Pain (VAS)	45	0.70	24.10	37	4.00	21.20
Hopman-Rock, 2000	Function (IRGL)	37	0.80	4.60	34	1.70	5.20
Baker, 2001	Pain (WOMAC)†	22	-79.00	88.00	22	-20.00	93.00
Baker, 2001	Function (WOMAC)†	22	-272.00	295.00	22	-119.00	323.00
Fransen, 2001	Pain (WOMAC)	83	-10.60	19.50	43	-1.50	19.40
Fransen, 2001	Function (WOMAC)	83	-7.70	19.90	43	-0.10	20.50
Halbert, 2001	Pain (WOMAC)	37	3.70	3.60	32	4.30	3.30
Halbert, 2001	Function (WOMAC)	37	11.40	10.70	32	13.80	10.20
Patrick, 2001	Pain (HAQ)	98	1.38	0.73	117	1.46	0.62
Patrick, 2001	Function (HAQ)	109	0.93	0.55	121	1.13	0.67
Belza, 2002	Pain (VAS)	101	-2.77	1.54	117	-1.46	0.62

n – Number of patients; Sd – Standard deviation; Negative mean estimate – Change score or change in the direction of reduce pain or function; † – Used 100mm Visual analogue scale.

Appendix 8d. continued

Author	Outcome	Treatment (n)	Treatment mean	Treatment sd	Control (n)	Control mean	Control sd
Thomas, 2002	Pain (WOMAC)	467	-1.27	3.60	316	-0.46	3.60
Thomas, 2002	Function (WOMAC)	467	-2.59	10.50	316	-0.02	10.50
Topp, 2002	Pain (WOMAC)	67	-1.53	3.20	35	-0.02	3.20
Topp, 2002	Function (WOMAC	67	-4.16	10.90	35	-0.17	10.90
Foley, 2003	Pain (WOMAC)	35	10.00	4.00	35	10.00	4.00
Foley, 2003	Function (WOMAC)	35	33.00	17.00	35	37.00	13.00
Quilty, 2003	Pain (VAS)	43	-7.93	27.50	44	-2.59	22.00
Quilty, 2003	Function (WOMAC)	43	-0.86	7.30	44	-0.27	7.60
Talbot, 2003	Pain (PRI)	17	-12.41	9.77	17	-10.12	4.64
Hughes, 2004	Pain (WOMAC)	68	4.90	3.40	43	6.20	4.30
Hughes, 2004	Function (WOMAC)	68	17.30	12.60	43	22.30	12.80
Keefe, 2004	Pain (AIMS)	16	3.19	1.85	18	4.03	2.08
Lin, 2004	Pain (WOMAC)	59	8.62	4.34	39	9.32	2.84
Lin, 2004	Function (WOMAC)	59	30.16	14.03	39	34.96	9.87
Messier, 2004	Pain (WOMAC)	80	0.42	3.50	78	1.06	3.40
Messier, 2004	Function (WOMAC)	80	3.07	11.60	78	3.40	11.50
Ravaud, 2004	Pain (VAS)	352	13.75	23.90	388	17.49	24.37
Bennell, 2005	Pain (VAS)	73	-2.20	1.70	67	-2.00	2.10
Bennell, 2005	Function (WOMAC)	73	7.80	8.70	67	8.20	10.00
Cochrane, 2005	Pain (WOMAC)	152	8.46	3.74	158	9.35	3.54
Cochrane, 2005	Function (WOMAC)	149	29.26	14.48	156	32.42	13.25
Tak, 2005	Pain (VAS)	35	-0.20	2.00	39	-0.05	2.40
Tak, 2005	Function (GARS)	23	-0.30	3.00	25	-0.20	2.70
Hay, 2006	Pain (WOMAC)	91	7.51	4.80	93	8.36	3.90
Hay, 2006	Function (WOMAC)	94	25.49	16.30	94	28.15	13.20

n – Number of patients; Sd – Standard deviation; Negative mean estimate – Change score or change in the direction of reduce pain or function.

Appendix 8d. continued

Author	Outcome	Treatment (n)	Treatment mean	Treatment sd	Control (n)	Control mean	Control sd
Mikesky, 2006	Pain (WOMAC)	15	-1.60	5.51	22	-0.36	3.44
Mikesky, 2006	Function (WOMAC)	15	-0.20	11.58	22	-1.93	9.11
Fransen, 2007	Pain (WOMAC)	55	27.30	18.70	41	40.00	16.20
Fransen, 2007	Function (WOMAC)	55	34.80	23.70	41	49.90	19.00
Hinman, 2007	Pain (WOMAC)†	36	143.00	79.00	35	198.00	108.00
Hinman, 2007	Function (WOMAC)†	36	598.00	316.00	35	656.00	373.00
Wang, 2007	Pain (VAS)	20	43.50	18.60	18	54.90	25.20
Wang, 2007	Function (HAQ)	20	0.90	0.40	18	1.00	0.50

n – Number of patients; Sd – Standard deviation; Negative mean estimate – Change score or change in the direction of reduce pain or function; † – Used 100mm Visual analogue scale.

Appendix 9. Criteria for judging risk of bias in the 'Risk of bias' assessment tool

RANDOM SEQUENCE GENERATION

Selection bias (biased allocation to interventions) due to inadequate generation of a randomised sequence.

Criteria for a judgement of 'Low risk' of bias.	The investigators describe a random component in the sequence generation process such as: • Referring to a random number table; • Using a computer random number generator; • Coin tossing; • Shuffling cards or envelopes; • Throwing dice; • Drawing of lots; • Minimization*. *Minimization may be implemented without a random element, and this is considered to be equivalent to being random.
Criteria for the judgement of 'High risk' of bias.	The investigators describe a non-random component in the sequence generation process. Usually, the description would involve some systematic, non-random approach, for example: • Sequence generated by odd or even date of birth; • Sequence generated by some rule based on date (or day) of admission; • Sequence generated by some rule based on hospital or clinic record number. Other non-random approaches happen much less frequently than the systematic approaches mentioned above and tend to be obvious. They usually involve judgement or some method of non-random categorization of participants, for example: • Allocation by judgement of the clinician; • Allocation by preference of the participant; • Allocation based on the results of a laboratory test or a series of tests; • Allocation by availability of the intervention.
Criteria for the judgement of 'Unclear risk' of bias.	Insufficient information about the sequence generation process to permit judgement of 'Low risk' or 'High risk'.

Appendix 9. continued

ALLOCATION CONCEALMENT	
Selection bias (biased allocation to interventions) due to inadequate concealment of allocations prior to assignment.	
Criteria for a judgement of 'Low risk' of bias.	Participants and investigators enrolling participants could not foresee assignment because one of the following, or an equivalent method, was used to conceal allocation: • Central allocation (including telephone, web-based and pharmacy-controlled randomization); • Sequentially numbered drug containers of identical appearance; • Sequentially numbered, opaque, sealed envelopes.
Criteria for the judgement of 'High risk' of bias.	Participants or investigators enrolling participants could possibly foresee assignments and thus introduce selection bias, such as allocation based on: • Using an open random allocation schedule (e.g. a list of random numbers); • Assignment envelopes were used without appropriate safeguards (e.g. if envelopes were unsealed or nonopaque or not sequentially numbered); • Alternation or rotation; • Date of birth; • Case record number; • Any other explicitly unconcealed procedure.
Criteria for the judgement of 'Unclear risk' of bias.	Insufficient information to permit judgement of 'Low risk' or 'High risk'. This is usually the case if the method of concealment is not described or not described in sufficient detail to allow a definite judgement – for example if the use of assignment envelopes is described, but it remains unclear whether envelopes were sequentially numbered, opaque and sealed.

Appendix 9. continued

BLINDING OF PARTICIPANTS AND PERSONNEL **Performance bias due to knowledge of the allocated interventions by participants and personnel during the study.**	
Criteria for a judgement of 'Low risk' of bias.	Any one of the following: • No blinding or incomplete blinding, but the review authors judge that the outcome is not likely to be influenced by lack of blinding; • Blinding of participants and key study personnel ensured, and unlikely that the blinding could have been broken.
Criteria for the judgement of 'High risk' of bias.	Any one of the following: • No blinding or incomplete blinding, and the outcome is likely to be influenced by lack of blinding; • Blinding of key study participants and personnel attempted, but likely that the blinding could have been broken, and the outcome is likely to be influenced by lack of blinding.
Criteria for the judgement of 'Unclear risk' of bias.	Any one of the following: • Insufficient information to permit judgement of 'Low risk' or 'High risk'; • The study did not address this outcome.
BLINDING OF OUTCOME ASSESSMENT **Detection bias due to knowledge of the allocated interventions by outcome assessors.**	
Criteria for a judgement of 'Low risk' of bias.	Any one of the following: • No blinding of outcome assessment, but the review authors judge that the outcome measurement is not likely to be influenced by lack of blinding; • Blinding of outcome assessment ensured, and unlikely that the blinding could have been broken.
Criteria for the judgement of 'High risk' of bias.	Any one of the following: • No blinding of outcome assessment, and the outcome measurement is likely to be influenced by lack of blinding; • Blinding of outcome assessment, but likely that the blinding could have been broken, and the outcome measurement is likely to be influenced by lack of blinding.
Criteria for the judgement of 'Unclear risk' of bias.	Any one of the following: • Insufficient information to permit judgement of 'Low risk' or 'High risk'; • The study did not address this outcome.

Appendix 9. continued

INCOMPLETE OUTCOME DATA **Attrition bias due to amount, nature or handling of incomplete outcome data.**	
Criteria for a judgement of 'Low risk' of bias.	Any one of the following: • No missing outcome data; • Reasons for missing outcome data unlikely to be related to true outcome (for survival data, censoring unlikely to be introducing bias); • Missing outcome data balanced in numbers across intervention groups, with similar reasons for missing data across groups; • For dichotomous outcome data, the proportion of missing outcomes compared with observed event risk not enough to have a clinically relevant impact on the intervention effect estimate; • For continuous outcome data, plausible effect size (difference in means or standardized difference in means) among missing outcomes not enough to have a clinically relevant impact on observed effect size; • Missing data have been imputed using appropriate methods.
Criteria for the judgement of 'High risk' of bias.	Any one of the following: • Reason for missing outcome data likely to be related to true outcome, with either imbalance in numbers or reasons for missing data across intervention groups; • For dichotomous outcome data, the proportion of missing outcomes compared with observed event risk enough to induce clinically relevant bias in intervention effect estimate; • For continuous outcome data, plausible effect size (difference in means or standardized difference in means) among missing outcomes enough to induce clinically relevant bias in observed effect size; • 'As-treated' analysis done with substantial departure of the intervention received from that assigned at randomization; • Potentially inappropriate application of simple imputation.
Criteria for the judgement of 'Unclear risk' of bias.	Any one of the following: • Insufficient reporting of attrition/exclusions to permit judgement of 'Low risk' or 'High risk' (e.g. number randomized not stated, no reasons for missing data provided); • The study did not address this outcome.

Appendix 9. continued

SELECTIVE REPORTING **Reporting bias due to selective outcome reporting.**	
Criteria for a judgement of 'Low risk' of bias.	Any of the following: • The study protocol is available and all of the study's pre-specified (primary and secondary) outcomes that are of interest in the review have been reported in the pre-specified way; • The study protocol is not available but it is clear that the published reports include all expected outcomes, including those that were pre-specified (convincing text of this nature may be uncommon).
Criteria for the judgement of 'High risk' of bias.	Any one of the following: • Not all of the study's pre-specified primary outcomes have been reported; • One or more primary outcomes is reported using measurements, analysis methods or subsets of the data (e.g. subscales) that were not pre-specified; • One or more reported primary outcomes were not pre-specified (unless clear justification for their reporting is provided, such as an unexpected adverse effect); • One or more outcomes of interest in the review are reported incompletely so that they cannot be entered in a meta-analysis; • The study report fails to include results for a key outcome that would be expected to have been reported for such a study.
Criteria for the judgement of 'Unclear risk' of bias.	Insufficient information to permit judgement of 'Low risk' or 'High risk'. It is likely that the majority of studies will fall into this category.

Appendix 9. continued

OTHER BIAS **Bias due to problems not covered elsewhere in the table.**	
Criteria for a judgement of 'Low risk' of bias.	The study appears to be free of other sources of bias.
Criteria for the judgement of 'High risk' of bias.	There is at least one important risk of bias. For example, the study: • Had a potential source of bias related to the specific study design used; or • Has been claimed to have been fraudulent; or • Had some other problem.
Criteria for the judgement of 'Unclear risk' of bias.	There may be a risk of bias, but there is either: • Insufficient information to assess whether an important risk of bias exists; or • Insufficient rationale or evidence that an identified problem will introduce bias.

Appendix 10. Usual care transition probabilities spread sheet

USUAL CARE SPREAD SHEET OF TRANSITION PROBABILITIES

p01	0.397824823	These are 3 months transition probabilities
p12	0.055128417	KEY: 0=No Pain, 1=Mild Pain,
p23	0.034996499	2=Moderate Pain, 3=Severe Pain
p10	0.061281037	
p21	0.0456221	
p32	0.046162135	

NB: Used arbitrary proportions for the cells

Three months

From / To	**No**	**Mild**	**Moderate**	**Severe**
No	0.602175177	0.397824823	0	0
Mild	0.061281037	0.883590545	0.055128417	0
Moderate	0	0.0456221	0.919381401	0.034996499
Severe	0	0	0.046162135	0.953837865

NB: Used sum products of rows and columns cell values of the 3 month transition probability matrix

Six months

From / To	**No**	**Mild**	**Moderate**	**Severe**
No	0.386994061	0.591074486	0.021931453	0
Mild	0.091049265	0.807626444	0.09939499	0.001929302
Moderate	0.00279577	0.082255367	0.849392747	0.065556117
Severe	0	0.002106014	0.086471801	0.911422185

NB: Used sum products of rows and columns cell values of the 6 month transition probability matrix

One year

From / To	**No**	**Mild**	**Moderate**	**Severe**
No	0.203642616	0.70791368	0.085865602	0.002578102
Mild	0.109047204	0.714257204	0.166863079	0.009832513
Moderate	0.010945943	0.138089291	0.735373881	0.115590885
Severe	0.000433506	0.010733109	0.152470166	0.836363218

NB: Used sum products of rows and columns cell values of the 1 year transition probability matrix

Two years

From / To	**No**	**Mild**	**Moderate**	**Severe**
No	0.11960732	0.661678631	0.199146956	0.019567092
Mild	0.101925145	0.610506899	0.252752475	0.034815481
Moderate	0.025386782	0.209167961	0.58238079	0.183064466
Severe	0.003290194	0.038004361	0.2414712	0.717234245

Appendix 10. continued

3 year observed proportions

From / To	No	Mild	Moderate	Severe
No	0.099	0.585	0.270	0.046
Mild	0.090	0.543	0.302	0.065
Moderate	0.034	0.250	0.493	0.223
Severe	0.008	0.071	0.294	0.628

Observed proportions at three years (crosstab B_line by 3yr pain groups)

	No	Mild	Moderate	Severe
No	0	0	0	0
Mild	0.09	0.54	0.30	0.07
Moder	0.03	0.25	0.49	0.22
Severe	0.01	0.07	0.29	0.63

Squares of errors for proportions

0.01	0.34	0.07	0.00
0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00

3 yrs predicted transitions numbers

From / To	No	Mild	Moderate	Severe	Start nos at baseline
No	0	0	0	0	0
Mild	77.59	467.94	259.85	55.62	861
Moderate	38.15	276.73	546.53	246.58	1108
Severe	5.31	48.23	200.79	429.66	684

Observed numbers at three years(crosstab B_line by 3yr pain groups)

	No	Mild	Moderate	Severe	Totals
No	0	0	0	0	0
Mild	77	467	258	59	861
Modera	37	278	547	246	1108
Severe	11	43	201	429	684

Squares of errors for number

0.00	0.00	0.00	0.00
0.34	0.89	3.42	11.42
1.33	1.61	0.22	0.34
32.34	27.39	0.04	0.44
Cells N	O	P	Q

KEY:

3 YRS OBSERVED PROPORTION = Guess Estimates

3 YRS PREDICTED = 3 Yrs Observed Proportion Cell Values * Start Numbers at B_line

SSE= (Predicted –Observed)^2

	SSE for numbers (Preferred method b'cos it gives appropriate weight)	
Sum of Squared Errors	=SUM(N45:Q48)	Used: DATA/TOOLS → Analysis →Solver → Equal to minimum → Δing
	79.78	$B$1:$B$6 → Solve

Appendix 11. Advice and information care transition probabilities spread sheet

Prob From/To				
p01	0.3978248		Relative probability	2.77067308
p12	0.0551284			
p23	0.0349965			
p10	0.1697897			
p21	0.1264039			
p32	0.1279002			

NB: Used arbitrary proportions for the cells

Three months

From / To	No	Mild	Moderate	Severe
No	0.6021752	0.397825	0	0
Mild	0.1697897	0.775082	0.055128417	0
Moderate	0	0.126404	0.838599576	0.0349965
Severe	0	0	0.127900186	0.87209981

NB: Used sum products of rows and columns cell values of the 3 months transition probability matrix

Six months

From / To	No	Mild	Moderate	Severe
No	0.4301615	0.547907	0.021931453	0
Mild	0.2338441	0.675267	0.088959704	0.0019293
Moderate	0.0214621	0.203976	0.714693756	0.05986849
Severe	0	0.016167	0.21879877	0.76503414

NB: Used sum products of rows and columns cell values of the 6 months transition probability matrix

One year

From / To	No	Mild	Moderate	Severe
No	0.3136344	0.610145	0.073849987	0.00237008
Mild	0.2604072	0.602287	0.129201159	0.00810466
Moderate	0.0722695	0.296245	0.542502627	0.08898261
Severe	0.0084765	0.067915	0.325200864	0.59840759

NB: Used sum products of rows and columns cell values of the 1 year transition probability matrix

Two years

From / To	No	Mild	Moderate	Severe
No	0.26261	0.580884	0.142827968	0.013678
Mild	0.2479185	0.560462	0.169774856	0.02184506
Moderate	0.1397712	0.389277	0.366858655	0.10409342
Severe	0.0489186	0.183056	0.380425682	0.38759938

NB: Used sum products of rows and columns cell values of the 1 year and 2 years transition probability matrices

Three years

From / To	No	Mild	Moderate	Severe
No	0.2440679	0.55333	0.17637728	0.0262245
Mild	0.2361587	0.540604	0.189928402	0.0333092
Moderate	0.1726025	0.435487	0.293490156	0.09842055
Severe	0.0937904	0.279123	0.359693315	0.26739323

Appendix 11. continued

	Mean WOMAC score within each pain group (μ)			
(μ)=	0	3	8	13

3 yrs predicted transition numbers with treatment

From / To	No	Mild	Modera	Severe	Start nos at baseline
No	0	0	0	0	0
Mild	203.33	465.46	163.53	28.68	861
Moderate	191.24	482.52	325.19	109.05	1108
Severe	64.15	190.92	246.03	182.90	684
Total(n)	458.73	1138.90	734.75	320.63	2653

3 yrs predicted/fitted transition numbers without treatment (i.e from usual care)

From / To	No	Mild	Moderate	Severe	
No	0	0	0	0	
Mild	77.59	467.94	259.85	55.62	
Moderate	38.15	276.73	546.53	246.58	
Severe	5.31	48.23	200.79	429.66	
Total(n)	121.05	792.91	1007.17	731.87	2653

NB: Predicted transitions numbers with treatment (6 months)

From / To	No	Mild	Modera	Severe	Start nos at baseline
No	0	0	0	0	0
Mild	201.34	581.40	76.59	1.66	861
Moderate	23.78	226.01	791.88	66.33	1108
Severe	0	11.06	149.66	523.28	684
Total(n)	225.12	818.47	1018.13	591.28	2653

6 Months Predicted transition numbers without treatm't (from usual care)

From / To	No	Mild	Moderate	Severe	
No	0	0	0	0	
Mild	78.39	695.37	85.58	1.66	
Moderate	3.10	91.14	941.13	72.64	
Severe	0	1.44	59.15	623.41	
Total(n)	81.49	787.95	1085.85	697.71	2653

sum (μ*n) in @ state	0	2455.4	8145.07	7686.6	18287.1	6.89	Overall mean	0	2363.8	8686.82	9070.2	20121	7.58	Overall mean
sum squared (μ*sum)	0	7366.21	65160.5	99926	172453	65.00	Overall mean sum square	0	7091.5	69494.6	117913	194499	73.31	Overall mean sum square
						17.49	Variance						15.79	Variance

Appendix 11. continued

Pooled variance	16.641
Pooled s.d.	4.0794
Stand mean diff	0.1694

Used: DATA/TOOLS → What if Analysis →Goal Seek → To Value=0.17(pooled estimate from SR/MA → Δing RR cell (E2) → Ok

Goal Seek	L67
To Value	0.17
Δing Cell	E2

Appendix 12. SF-6D algorithm for calculating utility scores – by Brazier and Roberts [2004]

```
*SF6D US Programme.*
*Deriving the SF-6D health state classification from the SF-12v2*.

*Revised by Donna Rowen in accordance with changes agreed by Qualitymetric Inc., John Brazier and
Dennis Fryback on 17th January 2007.
*Date: 27 February 2008

*Author: John Brazier
*Date: 24 June 2003

*Weighting of domain scores from Brazier JE, Roberts JR, (2004) The estimation of a preference-
based index from the SF-12. Medical Care, 42: 851-859.

*Please note the following before proceeding with the programme*
*SF-12 items are numbered 1-12.
*Programme uses raw (i.e.uncoded) SF-12 item responses.
*Designed for SF-12 version 2 (QM)
*The Health Institute, new England medical Centre, Boston, Massachusetts, March 1995).

*SF6D

*1. Physical functioning dimension.

IF (sf2=3) SFPhys = 1 .
IF (sf2=2) SFPhys = 2 .
IF (sf2=1) SFPhys=3 .
IF (sf2<1) OR (sf2>3) SFPhys = 9.
Execute.

*2. Role limitations dimension.

IF ((sf5=5) and (sf6=5)) SFRole = 1 .
IF ((sf5=1) OR (sf5=2) OR (sf5=3) OR (sf5=4)) AND (sf6=5) SFRole = 2 .
IF ((sf6=1) OR (sf6=2) OR (sf6=3) OR (sf6=4)) AND (sf5=5) SFRole = 3 .
IF ((sf5=1) OR (sf5=2) OR (sf5=3) OR (sf5=4)) AND ((sf6=1) OR (sf6=2) OR (sf6=3) OR (sf6=4)) SFRole =
4 .
IF ((sf5<1) OR (sf5>5)) AND ((sf6<1) OR (sf6>5)) SFRole = 9.
Execute .

*3. Social functioning dimension.
IF (sf12=5) SFSocial = 1 .
IF (sf12=4) SFSocial = 2 .
IF (sf12=3) SFSocial = 3 .
IF (sf12=2) SFSocial = 4 .
IF (sf12=1) SFSocial = 5 .
IF (sf12<1) OR (sf12>5) SFSocial = 9.
Execute.
```

Appendix 12. continued

```
*4. Bodily pain dimension.
IF (sf8=1) SFPain = 1 .
IF (sf8=2) SFPain = 2 .
IF (sf8=3) SFPain = 3 .
IF (sf8=4) SFPain = 4 .
IF (sf8=5) SFPain = 5 .
IF (sf8<1) OR (sf8>5) SFPain = 9.
Execute.

* 5. Mental health dimension.
IF (sf11=5) SFMental = 1 .
IF (sf11=4) SFMental=2 .
IF (sf11=3) SFMental=3 .
IF (sf11=2) SFMental=4 .
IF (sf11=1) SFMental=5 .
IF (sf11<1) OR (sf11>5) SFMental = 9.
Execute.

* 6. Vitality dimension.
IF (sf10=1) SFVital = 1 .
IF (sf10=2) SFVital = 2 .
IF (sf10=3) SFVital = 3 .
IF (sf10=4) SFVital = 4 .
IF (sf10=5) SFVital = 5 .
IF (sf10<1) OR (sf10>5) SFVital = 9.
Execute.

*Interaction term.
COMPUTE Most=0.
IF (SFPhys=3) Most=1.
IF (SFRole=3) Most =1.
IF (SFRole=4) Most=1.
IF (SFSocial=4) Most=1.
IF (SFSocial=5) Most=1.
IF (SFPain=4) Most=1.
IF (SFPain=5) Most=1.
IF (SFMental=4) Most=1.
IF (SFMental=5) Most=1.
IF (SFVital=4) Most=1.
IF (SFVital=5) Most=1.
Execute.

*Weighting of domain scores from Brazier JE, Roberts JR, (2004) The estimation of a preference-
based index from the SF-12. Medical Care, 42: 851-859.

IF (SFPhys=1) pf1=0 .
IF (SFPhys=2) pf1 = 0.
IF (SFPhys=3) pf1 = -.045 .
```

Appendix 12. continued

```
Execute.

IF (SFRole=1) rl1=0 .
IF (SFRole=2) rl1 = -.063 .
IF (SFRole=3) rl1 = -.063 .
IF (SFRole=4 ) rl1 = -.063 .

IF (SFSocial=1) sc1=0 .
IF (SFSocial=2) sc1=-.063 .
IF (SFSocial=3) sc1=-.066 .
IF (SFSocial=4) sc1=-.081 .
IF (SFsocial=5) sc1=-.093 .

IF (SFPain=1) pn1=0 .
IF (SFPain=2) pn1 = 0.
IF (SFPain=3) pn1 = -.042 .
IF (SFPain=4 ) pn1 = -.077 .
IF (SFPain=5) pn1 = -.137 .

IF (SFMental=1) mh1=0 .
IF (SFMental=2) mh1 = -.059 .
IF (SFMental=3) mh1 = -.059 .
IF (SFMental=4 ) mh1 = -.113 .
IF (SFMental=5) mh1 = -.134 .

IF (SFVital=1) v1 =0 .
IF (SFVital=2) v1 = - 0.078.
IF (SFVital=3) v1 = -.078 .
IF (SFVital=4 ) v1 = -.078 .
IF (SFVital=5) v1 = -.106.
Execute .

IF (Most=0) mst1=0 .
IF (most=1) mst1=-.077.
Execute.
Compute SF12ind = 1 + pf1+rl1+sc1+pn1+mh1+v1+mst1.
VARIABLE LABELS SF12ind  "SF-6D preference-based measured of health".
Execute .
```