Self-reported physical activity levels: measurement and assessment in community dwelling adults with or at risk of osteoarthritis.

Robert Smith

Ph.D.

June 2017

KEELE UNIVERSITY

Declaration

Degree for which thesis being submitted: Doctor of Philosophy

Title of thesis: Self-reported physical activity levels: measurement and assessment in community dwelling adults with or at risk of osteoarthritis.

This thesis contains confidential information and is subject to the protocol set down for the submission and examination of such a thesis.

NO

Date of submission: 27/03/2017 Original registration date: 22/08/2011

(Date of submission must comply with Regulation 2D)

Name of candidate: Robert Smith

Research Institute: Primary Care and Health Sciences

Name of Lead Supervisor: Dr. Emma Healey

I certify that:

(a) The thesis being submitted for examination is my own account of my own research

(b) My research has been conducted ethically. Where relevant a letter from the approving body confirming that ethical approval has been given has been bound in the thesis as an Annex

(c) The data and results presented are the genuine data and results actually obtained by me during the conduct of the research

(d) Where I have drawn on the work, ideas and results of others this has been appropriately acknowledged in the thesis

(e) Where any collaboration has taken place with one or more other researchers, I have included within an 'Acknowledgments' section in the thesis a clear statement of their contributions, in line with the relevant statement in the Code of Practice (see Note overleaf).

(f) The greater portion of the work described in the thesis has been undertaken subsequent to my registration for the higher degree for which I am submitting for examination

(g) Where part of the work described in the thesis has previously been incorporated in another thesis submitted by me for a higher degree (if any), this has been identified and acknowledged in the thesis

(h) The thesis submitted is within the required word limit as specified in the Regulations

Total words in submitted thesis (including text and footnotes, but excluding references and appendices): 63712

Signature of candidate 

Date 27/03/2017

Note

Extract from Code of Practice: If the research degree is set within a broader programme of work involving a group of investigators – particularly if this programme of work predates the candidate's registration – the candidate should provide an explicit statement (in an 'Acknowledgments' section) of the respective roles of the candidate and these other individuals in relevant aspects of the work reported in the thesis. For example, it should make clear, where relevant, the candidate's role in designing the study, developing data collection instruments, collecting primary data, analysing such data, and formulating conclusions from the analysis. Others involved in these aspects of the research should be named, and their contributions relative to that of the candidate should be specified (*this does not apply to the ordinary supervision, only if the supervisor or supervisory team has had greater than usual involvement*

Abstract

Physical activity (PA) is recommended for all adults with osteoarthritis (OA). Adults aged 45 years and over with joint pain, are likely to have already developed OA or are at risk of OA. This thesis examines self-reported PA levels in community dwelling adults aged 45 years and over with joint pain and how PA can be best measured within this population.

Self-reported PA levels of 14,212 adults (aged 45 years and over) with and without self-report joint pain demonstrated that adults aged 45 years and over with joint pain are less likely to be active compared to adults with no joint pain (OR= 0.75, 0.68-0.77 95%CI).

A systematic review appraised the measurement properties of twenty self-report PA instruments previously used in adult aged 45 years and over and OA or joint pain populations. The International Physical Activity Questionnaire short form (IPAQ-SF) and the Physical Activity Scale for the Elderly (PASE) appeared to be most suitable self-reported PA instruments in joint pain and OA research.

An analysis of measurement properties of the IPAQ-SF and PASE was conducted in 525 adults aged 45 years and over consulting primary care with joint pain. Reliability of the IPAQ-SF was lower (r=0.58, p=<0.01) compared to the PASE (ICC=0.69, 95%CI= 0.61-0.76, p=<0.001). Measurement error was large in the IPAQ-SF (-3942 to 4509 metabolic equivalents $(METS)^{-1minute\text{-}1week}$ 95% limit of agreement) and the PASE (-130.28 to 112.76 95% limit of agreement). In terms of construct validity, the IPAQ-SF and PASE correlated well with each other (r=0.62,

p=<0.01) and the SF-12 physical component score (PCS), (r=0.30, p=<0.01 & r=0.39, p=<0.01 respectively).

The implications of this thesis are that adults aged 45 years and over with joint pain are at higher risk of being inactive and that both the IPAQ-SF and PASE are poor in their measurement properties for measuring of PA in this population.

Declaration

The original research proposal and PhD were developed by Dr. Emma Healey and Professor Krysia Dziedzic as an application awarded for Acorn funding from Keele University. The PhD thesis plan, analysis, interpretation and discussion of findings on each portion of the PhD thesis are my own work.

The thesis contains secondary data analysis of a large population survey data set and a consultation questionnaire data set from The Management of Osteoarthritis in Consultations (MOSAICS) research project. The MOSAICS study was funded by the National Institute for Health Research (NIHR) in the 'Clinical osteoarthritis and joint pain in older people: optimal management in primary care' programme grant (Ref: RP-PG-0407-10386). My original contribution to the research of the MOSAICS project included taking telephone questions and inquiries from participants completing the survey, assisting the study statistician and custodian of the data in scoring of the physical activity questionnaires and data cleaning the population survey data set and consultation questionnaire data set.

The MOSAICS study is a complex mixed methods research project, which included a number of different research and data collection methods. The study design and outcome measure selection, including physical activity measures, had been finalised prior to the start of this PhD project. The findings of this thesis were made after the MOSAICS study had commenced. Therefore, there was limited opportunity for this PhD project to influence the design of the MOSAICS study.

The PhD candidate, however, did contribute to the MOSAICS study by conducting data cleaning, scoring and interpretation of physical activity measures in the

MOSAICS population survey and consultation questionnaires, working closely with the MOSAICS study statisticians. The candidate also regularly attended MOSAICS study meetings, and contributed to discussions regarding the physical activity aspect of the MOSAICS study in the data collection and analysis stages.

In this thesis I have planned and performed all analysis and written all chapters of the thesis with the support of my supervisory team (Dr. Emma Healey, Prof. Krysia Dziedzic, Prof. Gretl McHugh, Ebenezer Afolabi and Dr. Martyn Lewis).

Acknowledgments

I would like to thank my lead supervisor Dr. Emma Healey for all the work in proposing this PhD project, for her hard work and constructive feedback throughout the development of this thesis. I would also like to thank co-supervisors Professor Krysia Dziedzic and Professor Gretl McHugh for their contribution to the thesis.

A big thank you to my Dyslexia support tutors for their support, Thomas Stauton and Leanne Kemp. I would also like to thank Joanne Jordan, Ebenezer Afolabi and Dr. Martyn Lewis for their support, advice and feedback on methods and statistics. Thank you to Dr. Mel Holden for her work as a second reviewer of my systematic review. I also appreciate the support and help of Angela Pushpa-Rajah, Kris Clarkson and the rest of the MOSAICS project research team. I would also like to thank Sue Weir, Ashley Ford and Kelly Barnett.

Thank you to Keele University for funding my PhD and to the National Institute for Health Research (NIHR) for funding the MOSAICS project.

I would also like to thank the support and encouragement of my friends, Dr. April Woodward Dr. Tom Porter, Dr. Magda Rzewuska, Dr. Domenica Cotton, Cliona McRoberts, Kirsty Duncan, Daniel Green, Rebecca Lowey, Diane Smith, Tom Kingston and Jonathan Quicke.

Thank you to my mother, father, brother and grandparents for their support.

Finally, a big thank you to my girlfriend Maggie for her endless patience, support and encouragement.

PhD outputs

The work from this these has been published in two pieces of work. The results of the systematic review were published as an abstract in the Annals of Rheumatic Diseases. This research was also presented in a poster presentation at the European League Against Rhuematism (EULAR) 2013 conference.

The results of the physical activity component of the population survey were published as an abstract in the Journal of Rhuematology. This research was presented in a poster presentation at the British Soecity of Rheumatology 2014 conference.

Dedication

To my mother Jocelyn Smith and father David Smith.

Glossary

24PAR	24-hour Physical Activity Recall
AAP	The Adelaide Activities Profile
AAS	Active Australia Survey
ACR	American College of Rheumatology
ACSM	American College of Sports Medicine
AMED	Allied and Complementary Medicine
ANOVA	Analysis of Variance
ARS	Activity Rating Scale
ARUK	Athritis Research United Kingdom
AS	Ankylosing Spondylitis
BHF	British Heart Foundation
BMI	Body Mass Index
BNI	British Nursing Index

CI	Confidence Intervals
CINAHL	Cumulative Index of Nursing and Allied Health online
COSMIN	COnsensus-based Standards for the selection of health Measurement Instruments
CRD	Centre for Reviews and Dissemination reviews database
DALYS	Disablility Adjusted Life Years
DAQ	Daily Activity Questionnaire
DH	Department of Health
EQ-5D	EuroQoL 5 Dimensions
ES	Effect Size
EULAR	The European League Against Rheumatism
FDA	Food and Drug Administration
GPPAQ	General Practice Physical Activity Questionnaire
HAP	Human Activity Profile
HMIC	The Health Management Information Consortium

HR-PRO	Health Related Patient Reported Outcome
ICC	Intraclass Correlation
IMD	Index of Multiple Deprivation
IPAQ	International Physical Activity Questionnaire
IPAQ-SF	International Physical Activity Questionnaire Short Form
IPEQ	Incidental And Planned Activity Questionnaire For Older People
LASA	Longitudinal Ageing Study Amsterdam
LEAS	Lower-Extremity Activity Scale
MCS	Mental Component Score
METS	Metabolic equivalents
MLT-PAQ	Minnesota Leisure Time PA Questionnaire
MOSAICS	Management of OsteoArthritis In Consultations
MOST	Multicenter Osteoarthritis study
NICE	National Institute of Health and Care Excellence

NorStOP	The North Staffordshire Osteoarthritis Project
OA	Osteoarthrtis
OAI	The Osteoarthritis Initiative
OARSI	The Osteoarthritis Research Society International
OMERACT	Outcome Measures in Rheumatology
OR	Odd Ratio
P	P-value
PA	Physical Activity
PASE	Physical Activity Scales for the Elderly
PASIPD	PA Scale for Individuals with Physical Disablilities
PCS	Physical Component Score
PEDRO	Physiotherapy evidence database
PICOS	Population, Intervention, Comparison, Outcome, Setting
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses

PsycINFO	American Psychological Association PsycINFO
QAPAQ	Quality Assessment of Physical Activity Questionnaire
RCT	Randomised Control Trial
RR	Responsivness ratio
SDC	Smallest Detectable Change
SEM	Standardised Error of Measurement
SF-12	Short Form Health Survey 12 items
SF-36	Short Form Health Survey 36 items
SPSS	Statistical Package for the Social Sciences
SQUASH	Short Questionnaire To Assess Health Enhancing PA
SRM	Standardised Responsiveness Mean
STAR	Short Telephone Activity Rating
UCLAA	University of California, Los Angeles Activity
UK	United Kingdom

VAS	Visual Activity Scale
WoS	Web of Science
α	Alpha

Table of Contents

List of Tables

List of Figures

List of Boxes

Chapter one: Introduction

1.1 Definitions of physical activity, exercise and physical functioning

Physical activity (PA) is defined as “any physical bodily movement created by contractions of the skeletal muscle that raises metabolic rate of energy expenditure above that of resting” (Kaminsky, 2006). PA can include activities that are unplanned, occur incidentally and without structure, for example, walking to a location or physical work as part of a job (Bouchard et al., 2012). Exercise is a different concept under the umbrella of PA, and is defined as “a form of PA that is planned, structured and repetitively done to improve or maintain one or more attributes of physical fitness” (Kaminsky, 2006). Exercise differs from PA, as it is often done to specifically increase physical fitness, flexibility, or muscle strength (Bouchard et al., 2012). For example, conducting specific leg exercises, with a number of planned repetitions, can be a structured exercise used to increase muscle strength. This thesis considers PA in its broader definition, rather than only focussing on exercise. Other literature explored in this thesis has used PA and exercise interchangeably (Caspersen et al., 1985; Corbin et al., 2000; McArdle et al., 2010). The research in this thesis focusses consistently on PA, however when discussing other studies the term exercise has been used to reflect other studies’ use of this term. For example, an individual who goes for occasional walks, in the hope that it will help them to maintain some level of fitness, may consider this as PA, rather than exercise per se. cycling to work as a healthier mode of transport

could be interpreted as exercise or PA. In these examples, the terms PA and exercise overlap. In this thesis, the focus is on all forms of PA, including exercise and the related health outcomes (Blair et al., 1996).

Physical functioning is a separate concept to PA and exercise, and has been defined as “the activity limitations of an individual as a result of the individual’s physical health and the context of their environment” (Tomey & Sowers, 2009).

Physical functioning is however linked to PA and exercise, as higher levels of these have been associated with better physical functioning (Warburton et al., 2006). Physical functioning should be seen as an ability to complete physical tasks.

1.2 Measuring physical activity and exercise

The definition of PA links it to the body’s energy expenditure, measured in calories (kcals). However, measuring PA has proven to be difficult, as energy expenditure currently cannot be directly measured, nor can an individual accurately estimate their own calories of energy expired when performing a physical task. The most direct approach for measuring PA is the objective measurement of physiological parameters during physical activities that can estimate bodily energy expenditure (Ainsworth, 2009). The limitation of the most direct approaches to measuring energy expenditure is that the equipment required is often expensive, and impractical for measuring in daily life (Lamonte & Ainsworth, 2009). A favoured approach in population level research is subjective measurement of PA, using self-reporting instruments for individuals to recall and report their physical activity

participation over a period of time, retrospectively (Helmerhorst et al., 2012). More recently, some population level research has used objective measures such as accelerometers to measure physical activity. Details of the characteristics, including the strengths and limitations of objective and subjective approaches for measuring PA, is provided in Chapter Four (page 56).

There is a clear difference in how PA is defined when measured objectively and subjectively. Objective measures of PA estimate energy expenditure (Ainsworth, 2009), while subjective measuring involves individual's perceptions of how they define their own PA. These perceptions depend on their psychosocial perceptions of what constitutes meaningful PA, and in terms of time spent performing various physical activities (Melillo et al., 1996). This separates PA into two domains: objective measures physical energy expenditure over a given time period, and subjective measures social perceptions of an individual's PA levels. Systematic reviews of measuring subjective PA involve determination of validity by comparing the subjective measure to a 'gold standard' objective measure (Helmerhorst et al., 2012). Often, only moderate to weak correlations are found between objective and subjective PA approaches. This is unsurprising, as they can be viewed as measuring different domains of PA (Helmerhorst et al., 2012). In the discussion chapter of the thesis (Chapter Ten, page 308), the implications of selecting subjective or objective measures of PA are explored.

1.3 Why physical activity is important to measure

Increasing PA levels from a sedentary lifestyle is one of the most important improvements an adult could make for their health, as it has many physiological and psychological benefits (Powell et al., 2011). Regular PA can prevent many chronic health conditions and prolong life (Blair et al., 1996). A physically active lifestyle maintains good health and contributes towards the avoidance of health conditions (Haskell et al., 2007; Bull & Expert Working Groups, 2010). A lifestyle of long periods of inactivity or sedentary lifestyles are risk factors for many health conditions, poor health and factors that will decrease life expectancy (Rezende et al., 2014).

Due to the importance of PA, the UK Department of Health published guidelines for adults and adults aged 65 and over to attempt to be physically active for 150 minutes per week at moderate intensity, or 75 minutes at vigorous activity and to minimise the amount of sedentary time (Bull & Expert Working Groups, 2010). In the UK the current levels of PA among adults are low, with only a small percentage achieving recommended levels (Bull & Expert Working Groups, 2010). The percentage of adults achieving recommended levels further decreases with age and higher proportions of adults aged 65 years and over are completely sedentary compared with younger adults (BHF, 2012). Management of a wide range of chronic diseases such as cancer, cardiovascular disease, chronic obstructive pulmonary disease (COPD), diabetes, depression and musculoskeletal conditions can be improved through increased PA (Garcia-Aymerich et al., 2006; Harvey et al., 2010; Friedenreich et al., 2010 Savigny et al., 2012; Gillam & Steel, 2013; Jeon et al., 2013).

In developed and developing countries an observed increase in chronic disease, which is thought to be caused by increases in obesity and physical inactivity (World Health Organisation (WHO), 2002). As a result there has been great focus on increasing PA in adult populations (WHO, 2005) and given the importance of PA to health, it is important to measure PA in adults to address the health promoting issues of physical inactivity (Bauman et al., 2006). A valid and reliable measurement instrument of PA levels would allow for better understanding of PA. It could also be used to identify relationships between PA levels and health outcomes, be used to monitor a population's PA levels and can measure the impact of interventions aimed at increasing PA levels (Bauman et al., 2006).

1.4 Importance of physical activity in joint pain

OA is defined using the National Institute for Health and Care Excellence guidance (NICE, 2014); as a clinical syndrome of joint pain with some degree of limitation in physical function and reduced quality of life. A population where promoting regular PA is important is adults with joint pain, who are at risk of osteoarthritis (OA). A population based prevalence study of OA in UK adults aged 50 years and above estimated that 66.2% of the population have joint pain that interferes with their daily living, with 38.1% of all UK adults having disabling joint pain in at least one of the four regions of the body (Thomas et al., 2004). The joint pain and physical limitations individuals' experience affects the ability to perform daily tasks (NICE, 2014). Joint pain is defined as any pain in or around the joint. Joint pain in OA is frequently use-related and is often described as aching or throbbing, with episodes of sharp and stabbing pain on activity (Kidd, 2006; Lane et al., 2011). In adults aged 45 years and over, OA is the most common cause of joint pain and most

commonly affects the knees, hips, hands and feet (Hootman & Helmick, 2006; Bijlsma et al., 2011). OA becomes more prevalent in adults aged 45 years and over compared to individuals younger than 45 years (Cushnaghan & Dieppe, 1991).

OA is the most common chronic condition seen in primary care, where the majority of OA is managed and represents a large burden on primary care in the UK (Copper et al., 1998; Felson et al., 2007; ARUK, 2013). It is predicted that by 2030, OA will be the biggest cause of disability in the general population (Jagger et al., 2006). The focus and interest for this thesis is a population of adults aged 45 and over with peripheral joint pain; representing a population likely to have OA or at risk of developing OA. In the thesis the term target population will be used to refer to this population.

PA has an important role in reducing the burden of OA as there is sufficient evidence that aerobic fitness and muscle strengthening exercises are effective in reducing the joint pain symptoms described above and disability in adults with OA (Uthman et al., 2013). The evidence for and understanding of the role PA can play in developing OA and the management of OA symptoms has greatly increased in recent years (Uthman et al., 2013). There is now guidance from international organisations such as the European League Against Rheumatism (EULAR) (Fernandes et al., 2013) and the Osteoarthritis Research Society International (OARSI) (McAlindon et al., 2014), that recommends PA or exercise for all those with OA (NICE, 2014). To reduce the impact of disability caused in the UK, NICE published updated guidelines in 2014 for the management of OA in the knees,

hips, hands and feet (NICE, 2014). As part of these guidelines PA has been recommended as a core treatment for all adults with OA (NICE, 2014).

1.5 Measuring physical activity in joint pain

Despite the importance of PA in adults aged 45 years and over with joint pain being established, in the UK it is currently unknown what levels of PA adults aged 45 years and over with joint pain are routinely undertaking. It is also unknown how this population compares with the general population and in particular with adults aged 45 years and over who do not have joint pain. Studies in other countries indicate that adults with joint pain may be at higher risk of being inactive (de Groot et al., 2008; Holsgaard-Larsen & Roos, 2012) and so at risk of poorer health outcomes (Blair & Brodney, 1999). From guidelines, it is clear that PA is good for joint pain, the guidelines do not suggest how much PA is beneficial. Furthermore, a valid and reliable measure of PA is required to determine PA in this population. A systematic review of the measurement properties of PA instruments for adults with OA in the knee and hip has previously described that there was insufficient evidence for any instrument with adequate measurement properties for describing levels of PA (Terwee et al., 2011). Adults with joint pain can gain important health benefits from a physically active lifestyle, despite this, in the UK the PA levels of adults aged 45 years and over with joint pain has not been described. This is important as it can establish how many adults aged 45 years and over with joint pain are participating in PA and how many have a sedentary lifestyle. Describing PA levels in adults aged 45 years and over with joint pain also allows for relationships between PA levels and health outcomes in adults aged 45 years and over with joint pain to be investigated. A challenge in measuring PA in adults aged

45 years and over with joint pain is selecting an approach for making a measurement. There are two approaches that can be taken: objective or subjective. The objective approach includes using devices such as accelerometers, which have been shown to be a valid and reliable measure of PA in adults aged 45 years and over (Murphy, 2009). Subjective approaches are self-report outcome measures, where individuals are asked to report retrospectively their PA in which they have participated over a given length of time (Bauman et al., 2006). In this thesis, a subjective approach for measuring PA levels was chosen; objective measures are costly when utilised at a population level, cannot be used for all physical activities (Johannsen et al., 2010) and have a risk of low response rates without adequate data for analysis (Troiano et al., 2008). Self-report measures of PA levels are lower in cost compared to objective approaches and can cover all physical activities making them appropriate for measurement at a population level (Helmerhorst et al., 2012). There are a large number of self-report instruments available that measure PA, selecting the most appropriate will depend on the measurement properties, such as ease of use and precision in measurement, for adults aged 45 years and over with joint pain (De Vet et al., 2011). This thesis establishes the self-report levels of PA in adults aged 45 years and over with joint pain and provides an evaluation of the measurement properties of self-report PA instruments suitable for use in the target population.

1.6 Research question, aims and objectives

Considering the importance PA has for adults aged 45 years and over with joint pain, the current PA levels in this population in the UK are unknown and that it is not clear which self-report PA instrument is most appropriate. Therefore the

following research questions were addressed: What are the self-reported levels of PA in community dwelling adults aged 45 years and over with joint pain? What is the most appropriate method of measuring of self-reported levels of PA in the target population?

Three aims were constructed with a number of objectives to achieve each aim:

Aim one: To describe the self-reported levels of PA of UK community dwelling adults aged 45 years and over, with and without joint pain.

The three objectives addressing this aim were:

1a. To describe the overall levels of PA in an adult population aged 45 years and over, with and without self-reported joint pain.

1b. To describe the physical and mental health status of adults aged 45 years and over with self-reported joint pain reporting different levels of PA.

1c. To describe and compare levels of PA in two subgroups of adults aged 45 years and over with self-reported joint pain: in the lower limb only and in generalised joint pain (upper and lower limb).

Aim two: To evaluate the measurement properties of reproducible self-report PA instruments in adults aged 45 years and over with and without joint pain or OA.

The two objectives addressing this aim were:

1a. To identify self-report instruments using a systematic review of measures of PA previously used in adults aged 45 years and over with and without joint pain or OA.

1b. To describe the measurement properties of the self-report instruments identified by the systematic review measuring levels of PA in adults aged 45 years and over with and without joint pain or OA.

Aim three: To evaluate the measurement properties of the International PA Questionnaire Short Form (IPAQ-SF) and the PA Scale for the Elderly (PASE) in adults aged 45 years and over with joint pain.

The four objectives addressing aim three were in adults 45 years and over with joint pain and OA:

3a. To compare the responses of individual items and total scores between the IPAQ-SF and PASE.

3b. To assess and compare the reliability and measurement error of the IPAQ-SF and the PASE.

3c. To assess and compare the construct validity of total score and sub-domains in the IPAQ-SF and PASE.

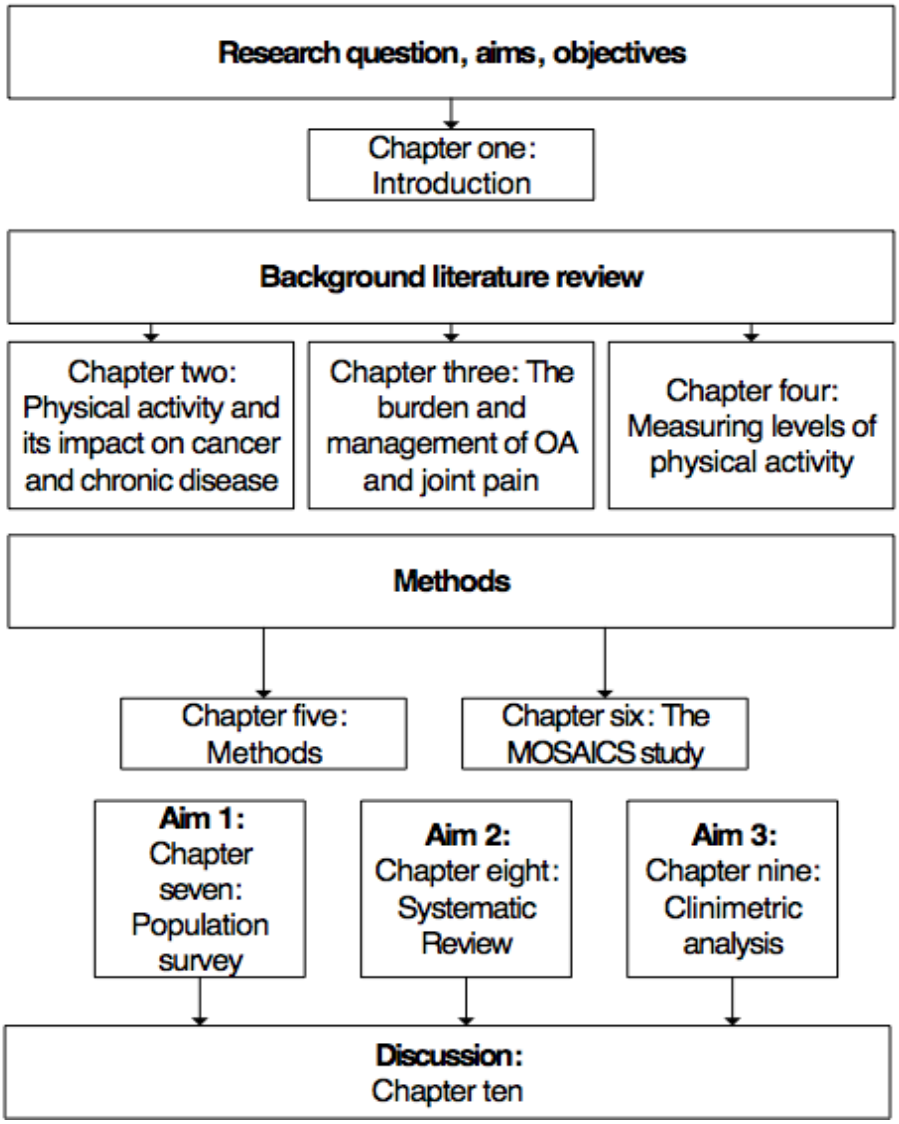
3d. To assess and compare the responsiveness of total score and sub-domains in the IPAQ-SF and PASE.

1.7 Summary

This introductory chapter provides an overview of the thesis together with its rationale: introducing and specifying the aims and objectives. The following chapters describe the background, methods, results and conclusion for each of the

three aims of the thesis; Figure 1.1 provides a summary of the overall structure of the thesis.

Figure 1.1 Format of thesis chapters



Chapter two: Physical activity and its impact on chronic disease

2.1 Introduction

This chapter provides an introduction to PA and its impact on chronic disease. This chapter will include:

- Definitions of PA and exercise
- PA recommendations
- Current levels of PA in the UK
- The burden of physical inactivity
- Physiological and psychological benefits of PA
- The impact of PA on common chronic diseases

2.2 Recommendations for physical activity

The most recent UK guidelines for PA were updated by an expert panel working group and published by the UK Department of Health (DH) in 2010 (Bull & Expert Working Groups, 2010). They are based on previously published UK PA guidelines from 2004 (DH, 2004), guidelines from other countries such as the United States (US Department of Health and Human Services, 2008) and Canada (Tremblay et al., 2011) and recent scientific evidence on PA (Haskell et al., 2007). The UK guidelines provide a uniform message for the type, duration, frequency

and intensity of PA recommended with those from other countries (Bull & Expert Working Groups, 2010).

The PA guidelines recommend that every adult aged 18-65 years should participate in at least 150 minutes of moderate aerobic activity per week (Bull & Expert Working Groups, 2010). This should be spread across the week, with at least 30 minutes of moderate aerobic PA on at least five days of the week, with each episode of activity needing to be continuous for at least 10 minutes to gain any health benefit (Bull & Expert Working Groups, 2010). Recommendations for vigorous intensity of PA state that: 75 minutes of vigorous aerobic PA per week is comparable to a benefit of 150 minutes of moderate aerobic PA (Rognmo et al., 2012). A combination of both moderate and vigorous activities can be used to achieve recommended guideline amounts (Bull & Expert Working Groups, 2010). This PA should be above and beyond that of usual daily activities (Kaminsky, 2006).

In addition to aerobic PA, it is recommended adults should participate in muscle strengthening activities on at least two days per week. Strengthening activities should use the major muscle groups including: upper and lower limb, core (around the abdominal muscles), the mid back and lower back muscles. This can include weight training using either weight training machines, free weights or weight bearing exercises, such as push-ups, sit ups or yoga poses (Bull & Expert Working Groups, 2010).

In terms of stretching, the guidelines suggest this may be beneficial for maintaining a full range of movement; they highlight a lack of adequate evidence for stating the required frequency and duration of stretching but recommend adults perform

stretching on two days of the week. These stretches should be performed in addition to warm up and cool down stretching, prior to and post PA (Bull & Expert Working Groups, 2010).

The importance of avoiding inactivity and minimising the amount of time spent sitting or sedentary was highlighted for all adults (Bull & Expert Working Groups, 2010). This formed part of the recommendations as previous research has demonstrated, participating in just low levels of PA, even below recommended amounts, provides some health benefits compared to being totally inactive (Blair et al., 1992; Warburton et al., 2006; Haskell et al., 2007). The health risks from total inactivity are high and reducing time spent sedentary, particularly sitting, is encouraged (Bull & Expert Working Groups, 2010). Table 2.1 provides a brief summary of the UK 2010 PA guidelines for adults.

Table 2.1 UK 2010 PA guidelines for adults

Activity Type	Recommendation
Moderate aerobic PA	At least 150 minutes of moderate intensity activity per week. Spread over 5 days a week, each about at least lasting 10 minutes
OR vigorous aerobic PA	At least 75 minutes vigorous intensity activities per week
OR a combination of moderate and vigorous PA	A combination of moderate and vigorous intensity equating to 150 minutes of moderate or 75 minutes of vigorous per week
AND strengthening activities	Strengthening exercises at least twice a week, incorporating all major muscle groups
AND stretching	Exercises that increase flexibility at least twice a week, this is in addition to a warm up and cool down exercises that accompany other activities

(Adapted from Bull & Expert Working Groups, 2010)

Elderly adults over 65 years of age have separate UK guidelines for PA levels to that of adults aged 18-65 years. These guidelines are identical in the type, duration and frequency of PA, strengthening exercises and flexibility exercises to that of adults aged 18-65 years, as shown in table 2.1. These guidelines do highlight that there is a need for PA to be individualised, with those participating in low levels to be encouraged to be more active even if this is still below recommended amounts. This also applies for adults with chronic illness, regardless of age. Activities should be tailored for individuals depending on their needs (Nelson et al., 2007).

Pedersen & Saltin (2006) conducted a systematic review of studies prescribing exercise therapy for adults with chronic conditions including an evaluation of the contraindications and precautions of exercise therapy for each condition. The systematic review concluded that PA was safe for adults with different chronic conditions as long as their symptoms and comorbidities were managed and stable. It also stated that the risk of poorer health due to not participating in PA in adults with chronic conditions was substantially higher compared to the safety risk of injury or worsening of symptoms associated with being physically active (Pedersen & Saltin, 2006).

2.3 Current levels of physical activity in UK

With the updated PA guidelines for adults and elderly adults over 65 years of age published in 2010, it has been possible to determine the proportion of the UK's population that are currently achieving recommended levels. A large population survey of adults and children was conducted in 2012 to examine PA levels of the UK's adult population (British Heart Foundation (BHF), 2012). The survey included

items on self-reported levels of PA in the UK. Table 2.2 summarises the data published in the BHF's report (BHF, 2012).

Table 2.2 Current levels of PA in UK adults

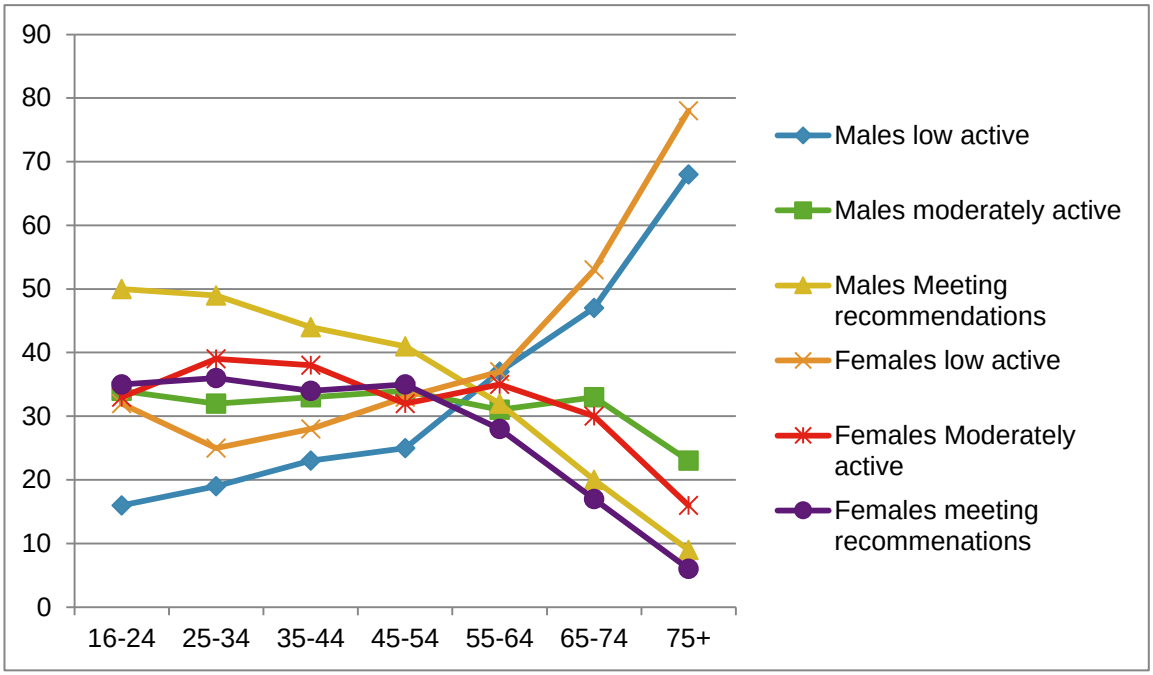
Gender	Age Group (years)	Low active (%)	Some Activity (%)	Meeting recommendations (%)
Males	16-24	16	30	50
	25-34	19	32	49
	35-44	23	33	44
	45-54	25	34	41
	55-64	37	31	32
	65-74	47	33	20
	75+	68	23	9
Females	16-24	32	33	35
	25-34	25	39	36
	35-44	28	39	34
	45-54	33	32	35
	55-64	37	34	28
	65-74	53	30	17
	75+	78	16	6

(Adapted from BHF, 2012) Key- meeting recommendations: 30 minutes or more of moderate or vigorous activity on at least 5 days a week; Some activity: 30 minutes or more of moderate or vigorous activity on 1 to 4 days a week; Low active: lower levels of activity.

Levels of PA in females was lower in all age groups compared with males, with smaller proportions of females meeting recommended levels of PA and higher proportions of females reporting low levels of activity for all age categories compared to males (table 2.2). Lowering of levels of activity can be seen with ageing in males and females, with elderly adults (65 years and over) reporting lower levels of activity. The same trends were also shown in females across age groups with increasing proportions reporting being inactive and decreasing

proportions meeting recommended levels as age increased. Figure 2.1 shows the relationship of ageing and levels of PA in male and female adults (BHF, 2012). The lines clearly show increased low active levels and reduced proportions of meeting recommendations in males and females with age increasing.

Figure 2.1 Levels of PA as ageing increases in UK male and female adults



(Adapted from BHF, 2012)

The report also highlighted that compared with previous data from 2003, levels of inactivity had reduced and more people were meeting recommended levels of activity across age groups for both males and females (BHF, 2012). While this suggests that the situation in the UK is improving because of an increase in PA level between 2003 and 2012, the data only considers population percentages rather than actual figures of the population. With the percentage of ageing adults increasing in the UK, the overall levels of PA are potentially set to reduce further (BHF, 2012; Christensen et al., 2009). This change in population-age demographic

will likely cause a lowering in the UK's population PA levels if the current trends in ageing demographics and PA levels continue (BHF, 2012; Christensen et al., 2009). There has been further research not only focusing on the importance of being physically active but also on avoiding sedentary lifestyles. Table 2.3 displays the objectively measured average distribution of time spent active and inactive during an average adult's (mean age 53.4 years) waking hours in a day. This was taken from a survey study of 173 adults (Healy et al., 2007).

Table 2.3 Average distribution of hours per day spent sedentary and active during adults' waking hours

Activity	Hours per day
Sedentary	9.3
Light intensity PA (<3.0 METS)	6.5
Moderate to vigorous intensity PA (≥3.0 METS)	0.7

Table 2.3 shows that average adults spend the largest proportion of their waking day sitting or sedentary (Healy et al., 2007). Increased amounts of time spent sedentary is an independent risk factor for developing metabolic syndrome (Wijndaele et al., 2009). Metabolic syndrome is a term for a group of risk factors associated with risk of cardiovascular diseases and diabetes (Isomaa et al., 2001). Risk factors included in metabolic syndrome are hypertension, high blood cholesterol, high blood triglycerides and high plasma glucose. Sedentary behaviour has been shown to be a predictor for weight-gain in Australian adult females even after adjusting for diet and PA behaviour (Brown et al., 2005). Reducing time spent sedentary or breaking up prolonged time spent sedentary has also been associated with a reduced risk of metabolic syndrome (Healy et al., 2008).

2.4 The burden of physical inactivity

With the absolute levels of PA projected to reduce in the future, the implications of this for society and the public need to be considered. Recently, the effects of being physically inactive have received much attention in the public health setting (Lee et al., 2012). Craig et al. (2012) argue that inactivity should be public health's biggest priority due to: its high prevalence and association with numerous chronic health conditions, the likelihood of inactivity causing a health condition and the impact of such health conditions. Given the impact physical inactivity has on public health, targeting reduced inactivity is a high priority (Craig et al., 2012).

Physical inactivity has been reported by the World Health Organisation (WHO) to be the fourth leading risk factor for mortality in the world and alone it accounted for six per cent of all deaths in 2010 (WHO, 2010). This figure could be much higher given the number of other leading risk factors associated with inactivity, such as cardiovascular disease (WHO, 2010). Das & Horton (2012) noted that despite convincing evidence for the health benefit of PA for the whole population existing since the 1950s, emphasis on increasing levels of PA in society has been low with advice to be more active only given to individuals rather than the whole population.

It is estimated in the UK that the cost of physical inactivity to the National Health Service (NHS) was £936 million between 2006 and 2007 (Allender et al., 2007). Obesity which can be caused by inactivity, along with poor diet cost the NHS a further £5.1 billion (Scarborough et al., 2011). Further estimates of the burden of physical inactivity suggest it is independently responsible for three per cent of all disability adjusted life years lost in the UK with this percentage most likely

increasing when taking into consideration other risk factors such as obesity and ischemic heart disease (Allender et al., 2007).

2.5 Physiological benefits of physical activity

There is considerable evidence for PA in the prevention and treatment of many different diseases and health conditions (Blair et al., 1996; Powell et al., 2011). In adults who participate in PA on a regular basis there is an association in increases in overall quality of life, reduced mortality and increased lifespan (Blair et al., 1989; Wen et al., 2011).

Blair et al. (1989) conducted a longitudinal study in the United States investigating low cardiorespiratory fitness as a risk factor for all-cause mortality compared to other risk factors in an adult population. Attributable fraction is the incidence of a disease that would have been prevented if a risk factor exposure was eliminated. In the case of this study, attributable fraction was calculated to determine the percentage of deaths in a year that could have been avoided if the risk factor, low cardiorespiratory fitness, did not exist (Greenland et al., 1988). Blair et al. (1989) found cardiorespiratory fitness was shown to have the highest attributable fraction compared to hypertension, smoking, high cholesterol, diabetes and obesity, in both males and females. This shows the substantial health benefits of improving physical fitness by being more physically active. The importance and impact of PA has on health would seem to be a leading risk factor for poor health and therefore important to public health policy making (Blair et al., 1989).

A large prospective cohort study conducted in Taiwan of 416,175 individuals examined the dose-response relationship of PA and mortality in all adults aged 20

years and over (Wen et al., 2011). This study found being physically active even at low levels was shown to reduce all-cause mortality by 14 per cent, compared to no activity at all, while meeting the minimum recommended levels of PA reduced mortality by 20 per cent. High levels of PA beyond the recommended guidelines reduced mortality by 29 per cent and very high levels of PA reduced mortality by 35 per cent (Wen et al., 2011).

This reduction in all-cause mortality is due to the physiological benefit of PA (Blair et al., 1996). The physiological benefits of PA are triggered by functional improvements and biochemical changes as a response to frequent bouts of PA (Garber et al., 2011). Functional improvements include a greater metabolic capacity and increases in nutrients supplied to the skeletal muscle by the blood supply (Fentem, 1994). The greater metabolic capacity occurs within the skeletal muscle fibres which includes muscle fibre recruitment and adaptions within muscle cells that leads to increased stamina (Holloszy & Coyle, 1984). Physiological responses to PA include: increased density of muscle cell mitochondria, bone density, oxidative fibre type, processing fatty acid, blood vessel recruitment and increased cardiac stroke volume (Holloszy & Coyle, 1984; Hamilton & Booth, 2000; Vina et al., 2012). Increased nutrients to the skeletal muscles from the blood supply come from changes within the cardiac system, increasing the stroke volume to increase blood supply and regulation of arterial blood pressure and from a denser network of peripheral blood vessels to the skeletal muscles (McArdle et al., 2010). There are a wide variety of biochemical changes that occur as a response to regular PA, the most significant changes to physiological benefits are controlling of body weight (Slentz et al., 2004), maintenance of bone mass (Morris et al., 1997), regulation of body fat (Slentz et al., 2004), regulation of blood

glucose insulin sensitive (Goodyear et al., 1998), inhibiting blood clots and lowering blood pressure (Fentem, 1994). Adults that participate in regular PA have also been shown to have higher anti-inflammatory markers compared to sedentary adults (Gleeson et al., 2011). This is important as inflammatory markers have shown to be linked to poorer health (Handschin & Spiegelman, 2008).

2.6 Psychological benefits of physical activity

Along with physiological benefits, regular PA has positive psychological benefits, which have been shown to have long term effects on: depression (Mead et al., 2009), anxiety (Petruzzello et al., 1991) and vigour in adults (DiLorenzo et al., 1999). A Finnish cross sectional study found associations between PA and depression, anger, social distrust, stress, social integration, perceived health and perceived fitness with the greatest psychological benefits associated with higher frequency of PA (Hassmen et al., 2000). A study on the dose response of PA to psychological benefits showed that levels of PA comparable to guideline levels (Bull & expert working panel, 2010) had the greatest impact, while lower PA levels were equal in impact to placebo (Dunn et al., 2005). These associations in psychological benefit are backed up by evidence of PA causing biochemical changes within the central nervous system; PA is associated with release of endorphin hormones and dopamine chemicals that have positive impact on mood within the central nervous system (Craft & Perna, 2004).

A systematic review has also identified a strong relationship between high levels of PA and higher levels of health-related quality of life (Bize et al., 2007). This association of PA and health-related quality of life has been shown in a number of randomised control trials (RCTs) (Ashley et al., 2001; Brand et al., 2005; Patonen

et al., 1998; Stewart et al., 1993), cohort studies (Malmberg et al., 2005; Tessier et al., 2007; Wendel-Vos et al., 2004) and cross sectional studies (Brown et al., 2003; Daskapan et al., 2005; LaForge et al., 1999; Lindholm et al., 2003; Riise et al., 2003; Vuillemin et al., 2005). These studies used two different methods to measure PA: observed PA sessions (Ashley et al., 2001; Brand et al., 2005; Patonen et al., 1998; Stewart et al., 1993) or self-report PA instruments (Brown et al., 2003; Daskapan et al., 2005; LaForge et al., 1999; Malmberg et al., 2005; Lindholm et al., 2003; Riise et al., 2003; Tessier et al., 2007; Vuillemin et al., 2005; Wendel-Vos et al., 2004). Various methods were used to measure health-related quality of life, these included a measure of generic self-report health-related quality of life using the SF-36 and reporting number of days without disability or reporting number of unhealthy days in previous 30 days (Bize et al., 2007).

2.7 The impact of physical activity on cancer and common chronic diseases

Populations that could gain most from the health benefits of regular PA are those with pre-existing health conditions and those at risk of developing common chronic conditions (Bull & Expert Working Groups, 2010). In adults with cancer or common chronic conditions participating in PA, this could slow or halt disease progression, helping management of symptoms and reduce the risk of developing other morbidities (Nelson et al., 2007). Due to the symptoms of their health conditions or disease, individuals with chronic conditions can find it difficult to be physically active (Nelson et al., 2007).

Being physically active is recommended not only to prevent the risk of developing other chronic conditions but because it has also been shown to have a therapeutic

effect in reducing or preventing symptoms, as well as slowing or halting the rate of disease pathology (Nelson et al., 2007).

(1) Cancer

PA has an important role in the prevention and treatment of many different types of cancer (Lee, 2003; Schmitz et al., 2005). A meta-analysis showed there is sufficiently strong evidence for recommended levels of PA in adults to reduce the risk of developing colon and treating fatigue symptoms, improving health-related quality of life in cancer treatment patients for cancer in (Galvao & Newton, 2005). PA has also been shown to be important for improving outcomes for breast cancer (Mock et al., 1997), ovarian cancer (Stevinson et al., 2009), lung cancer (Knols et al., 2005) and prostate cancer (Segal et al., 2009). These finds were also summaried in a systematic review on the effect of PA in different cancers (Friedenreich et al., 2010). PA has shown to be an important part of therapy in those who have survived cancer. The American Cancer Society (ACR) recommends PA as a treatment in their latest guidelines along with nutritional guidance for the management of cancer survivors (Rock et al., 2012). The American College of Sports Medicine (ACSM) has also published guidelines on the prescription of PA as part of self-management in those at the post-treatment survivorship stage of cancer (Nelson et al., 2007). Cancer survivies are recommended to avoid long periods of inactivity (Schmitz et al., 2010).

(2) Cardiovascular disease

There are seven clinical conditions that are connected to the heart or the vascular system which can be effectively prevented or treated with PA. Meta-analysis of RCTs have identified conditions which can be treated or prevented with PA include: coronary heart disease (Taylor et al., 2004), heart failure (Piepoli et al., 2004), hypertension (Whelton et al., 2002) and stroke (Kwakkel et al., 2004). All of these conditions have epidemiological evidence that PA is a protective factor against their development (Gerber et al., 2011; Petersen et al., 2012). A meta-analysis of RCTs has demonstrated that increasing levels of PA from sedentary levels decreases the risk of coronary heart disease in both males and females (Sattelmair et al., 2011). Although frequent exertions of vigorous intensity PA at elite endurance training levels has shown to be a risk factor in developing atrial fibrillation (Pelliccia et al., 2005), more regular levels of light and moderate PA in regular daily life have been shown to reduce incidence of atrial fibrillation (Mozaffarian et al., 2008). PA based cardiac rehabilitation is recommended for those who have angina, have had a myocardial infarction or have undergone a coronary heart bypass. Cardiac rehabilitation, which typically includes PA interventions as a core treatment is effective in preventing disability in adult populations and preventing subsequent cardiovascular events or death for cardiac causes (Giannuzzi et al., 2003). New developments in clinical care and cardiac rehabilitation has been credited with improving survival in patients with cardiac disease or events and health-related quality of life through reduced disability (Balady et al., 2007). Individuals attending PA based cardiac rehabilitation demonstrated improved survival rates and reduced risk of a myocardial infarction

compared to those who did not (Fletcher et al., 2013; Giacomantonio et al., 2013 Hammill et al., 2010).

(3) Diabetes mellitus (Type II diabetes)

A systematic review of prospective cohort studies investigated the risk of developing type II diabetes in those that undertook regular moderate PA in comparison to those who were sedentary (Jeon et al., 2007). They found that adults who had participated in regular moderate PA had a lower risk of developing type II diabetes compared with sedentary adults or adults that did very little PA (Jeon et al., 2007). A meta-analysis of RCTs investigating adults with type II diabetes found PA interventions lasting eight weeks or greater increased glycaemic control independent of changes to body weight compared to a non-PA control group (Sigal et al., 2006). Both aerobic and strengthening resistance training has been shown to result in clinically significant reduction in glycosylated haemoglobin (HbA1c). A clinically significant reduction in HbA1c signifies a reduced risk of complications to diabetes connected with hyperglycaemia causing damage to organs and other tissues, allowing for management of diabetes (Boule et al., 2001)

(4) Chronic obstructive pulmonary disease (COPD)

PA based pulmonary rehabilitation for adults with chronic obstructive pulmonary disease (COPD) have been shown not only to improve cardiorespiratory physical fitness (Watz et al., 2009) but also reduce the number of hospital admissions for an exacerbation of COPD and reduce the incidence of mortality connected with complications in COPD (Garcia-Aymerich et al., 2006). PA does not improve the

functional capacity of the lungs that is often affected by COPD; instead PA reduces the physical disability that develops with COPD (Pitta et al., 2005). In adults with COPD, PA improves aerobic fitness and endurance to perform activities such as walking for longer periods before becoming breathless (Effing et al., 2011). Regular PA targeting at increasing aerobic fitness and strengthening exercises to increase musculoskeletal functioning are recommended for adults with COPD by the American Thoracic Society and the European Respiratory Society (ERS) (Pauwels et al., 2012). In the UK, NICE recommends multi-disciplinary pulmonary rehabilitation with PA as a core intervention for adults with COPD. The aim of which is to reduce disability and increase independence in those with COPD symptoms (Halpin, 2004).

(5) Mental health

The term mental health conditions cover a wide number of different psychological conditions: stress, depression, anxiety, Alzheimer's disease and dementia (Deslandes et al., 2009). NICE recommend PA for all adults aged 45 years and over in primary care and residential care to promote mental wellbeing (NICE, 2008). It is recommended that adults aged 45 years and over participate in guided sessions with mixed range of moderate intensity physical activities on at least one or two sessions a week (NICE, 2008). These sessions should be accompanied by further PA during the week matching that to PA guidelines for adults aged 45 years and over (Bull & expert working panel, 2010). A review of the literature on PA interventions for anxiety and depression shows favourable outcomes in RCTs and the reduced risk developing of the psychological conditions depression and anxiety in adults (Harvey et al., 2010; Carek et al., 2011). Those who increase

their PA levels to meet the recommended levels showed larger reductions in depression compared to those achieving lower levels of PA (Harvey et al., 2010). This is thought to be due to a number of neurophysiological mechanisms within the central nervous system, which improve mood, self-satisfaction and perceptions of well-being (Carek et al., 2011). Adults who self-reported being physically active during the last thirty days were found to have a lower risk of diagnosis of lifetime depression or anxiety (Strine et al., 2008). Interestingly a study has demonstrated that lower intensity PA showed similar reductions in depression compared to a control group who received stretching exercises only (Martinsen, 2008). Secondary analysis of a population based cohort study found adults aged 45 years and over aged 65-79 years who participated in PA at least twice a week were at a reduced risk of developing dementia and Alzheimer's disease later in life compared to those who were sedentary (Rovio et al., 2005). Longitudinal prospective studies have linked increased volumes of weekly PA as a protective factor in preventing cognitive decline, a key factor for development of Alzheimer's disease and dementia, in adults aged 45 years and over (Laurin et al., 2001; Lytle et al., 2004). An RCT compared a PA intervention to a non-PA control in adults aged 45 years and over with some signs of cognitive memory problems, the PA intervention showed a modest but significant improvement in subjective cognitive function compared to controls at 18 months (Lautenschlager et al., 2008). PA also appears to be effective in treating elderly adults in nursing homes already with a diagnosis of Alzheimer's disease or dementia in reducing decline of disability (Rolland et al., 2007).

(6) Musculoskeletal conditions

PA is a widely accepted and used treatment for low back pain, a common site of musculoskeletal pain among adults (Hayden et al., 2005). NICE recommend exercise therapy in their latest guidelines for treating lower back pain (Savigny et al., 2009), as it has been proven to reduce pain intensity, disability (Chou & Huffman, 2007) and promotes faster recovery and return to work (Waddell & Burton, 2001).

A Cochrane systematic review on the treatment of fibromyalgia showed that aerobic exercises have a positive effect on general well-being, fibromyalgia pain, tender points and can reduce depression in this population (Busch et al., 2008). As a consequence of the evidence for aerobic exercises benefiting fibromyalgia, the European League Against Rheumatism (EULAR) have recommended exercise based interventions for the management of fibromyalgia symptoms (Carville et al., 2008).

In postmenopausal women who are at high risk of osteoporosis, a systematic review showed PA can reduce bone loss and may increase bone mineral density (Schmitt et al., 2009). The UK guidelines for PA in elderly adults recommend increasing PA to reduce risk of losing bone mineral density or fractures (Bull & expert working group, 2010). A large meta-analysis showed that PA is associated with improved musculoskeletal health and reduces the risk of fractures in those at risk of osteoporosis (Qu et al., 2013).

Exercise therapy is used as part of self-management in patients with ankylosing spondylitis (AS) to improve musculoskeletal mobility, strength and aerobic fitness

(Hidding et al., 1993; Lim et al., 2005; Ince et al., 2006). Exercise therapy that includes flexibility, strengthening and aerobic activities improves joint range of motion, aerobic fitness, self-reported pain scores and physical functioning in patients with AS (Hidding et al., 1993; Analay et al., 2003; Dagfinrud et al., 2005). In a joint statement, both the international society for AS and the EULAR recommend exercise interventions as part of managing AS (Zochling et al., 2006).

2.8 Summary

This chapter has described the PA guidelines for adults, elderly adults and adults with chronic diseases in the UK. The chapter also displayed levels of PA in the UK for adults, showing the most recent population data on levels of PA and predictions for overall reduction in levels of PA in future years. This chapter has also introduced the impact PA has on common chronic diseases.

Chapter three: The burden and management of OA and joint pain

3.1 Introduction

Chapter two described the psychological and physiological benefits of PA and the impact it has on chronic diseases. This chapter further explores the importance of PA in joint pain or osteoarthritis (OA). This chapter will include:

- A description of joint pain and OA populations
- The burden of OA in the UK public health
- The risk factors for OA and why it is highly prevalent among UK adults aged 45 years and over
- The pathology of OA
- The diagnosis and symptoms of OA
- How PA and exercise benefits in the management of OA symptoms
- Levels of PA in adults with OA.

3.2 Definition of joint pain and osteoarthritis populations

Musculoskeletal joint pain is highly prevalent in adults aged 45 years and over (Thomas et al., 2004; Litwic et al., 2013). In a UK population of adults 75 years and over 83 percent reported some degree of joint pain over a year (Donald & Foy, 2004). OA is the most common cause of musculoskeletal joint pain in adults aged 45 years and over with prevalence increasing with ageing; the knee, hip,

hand and foot are the most commonly affected sites (Hootman & Helmick, 2006; Biljlsma et al., 2011). It is likely that adults aged 45 and over with symptomatic joint pain have either already developed or are at risk of developing OA. This is because ageing is an independent risk factor for developing OA and joint pain is a symptom of OA (Lane et al., 2011). NICE define OA as a clinical syndrome of joint pain in adults aged 45 years and over.

3.3 Burden of joint pain and osteoarthritis in the UK public health

Joint pain and OA are important due to their prevalence and burden to public health in the UK (Litwic et al., 2013). In the UK the prevalence of musculoskeletal pain reported in a cross-sectional survey of adults in a general population aged 50 years and over was 66.2%, with a musculoskeletal pain that interferes with daily life prevalence of 38.1% in all responders, which equated to 58.7% of responders with musculoskeletal pain (Thomas et al., 2004). When separated into anatomical regions, the prevalence of pain in the knee ranged from 35.4% increasing with age to 37.7%, hips from 25.6% to 28.3%, feet from 21.1% to 23.5% and hands from 16.9% to 25.6% (Thomas et al., 2004).

A meta-analysis of studies from different countries across the globe showed an overall prevalence of OA in the knee of 23.9%, 10.9% in the hips and 43.3% in the hands (Pereira et al., 2011). These percentages take into account studies examining the prevalence of radiographic, self-reported and symptomatic OA together. In the UK it is estimated that 6.65 million adults suffer from OA (Oxford Economics, 2010). Musculoskeletal pain, including OA account for the fifth most years lost to disability in all UK adults behind low back pain, falls, depression and neck pain (Murray et al., 2013). In 2010, OA was the second largest cause of

disability adjusted life years of musculoskeletal disorders, after lower back and neck pain (Murray et al., 2013).

The majority of the adult population within the UK will access primary care for the management of their OA symptoms (Peat et al., 2001). Age and gender standardised annual consultation prevalence per 10,000 in the UK for musculoskeletal problems in or around the knee was 324, 208 for in or around the foot, 132 for in or around the hand and 115 for in or around the hip (Jordan et al., 2010). The prevalence of all OA primary care consultations when standardised for age and gender was 426 per 10,000 (Jordan et al., 2007). The prevalence of OA increased with age in the UK, the estimated number of individuals who sought treatment for their OA symptoms decreased from 4.33 million aged 45-64 years to 2.15 million 65-74 years and remained stable at 2.27 million in elderly adults 75 years and over. This was recorded over seven years of primary care consultation data and suggested that one in three adults aged 45 years and over have sought treatment for OA (Arthritis Research UK (ARUK), 2013).

In the UK, the estimated total direct cost of health care, including cost of pharmacological treatment, for OA is £783 per person per year (Oxford Economics, 2010), with the total direct cost of OA in the UK estimated to be £5.2 billion (Oxford Economics, 2010). It is also estimated that £7.1 billion is lost in the value of healthy life lost, this is based on annual disability adjusted life years (DALYs) lost (Oxford Economics, 2010). A systematic review of economic costs of OA globally and in the UK, observed a rise in economic burden of disease, in both direct and indirect cost (Chen et al., 2012). In economic costs this represents a

large burden on UK primary care to manage adults aged 45 years and over with joint pain symptoms.

3.4 The risk factors for developing joint pain or osteoarthritis

Risks factors for OA in the knee, hip, hand and foot include age, female gender, genetics, obesity, Heberden's nodes, repetitive occupational activities / injury, increased joint loading, joint malalignment or imbalance of joint muscle strength and weakness (Arden & Nevitt, 2006; Blagojevic et al., 2010; Pereira et al., 2011; Loeser, 2013; Neogi & Zhang, 2013). These factors act as risks by increasing the susceptibility of joint injury, direct damage of the joint or by acting as a hindrance to the repair of the damaged tissue.

3.4.1 Age

With the UK population demographics moving towards an ageing population, the risk for developing OA is set to increase (Loeser, 2013). Felson et al. (1987) observed ageing as a risk factor for developing radiographic diagnosis of knee OA when the prevalence of OA was found to increase in the Framingham study cohort above the age of 65 years. Ageing appears to have a linear association with the increase in prevalence of developing OA, particularly in the knee and hand after 45 years, with the foot and hip increasing in prevalence after 50 years old (Oliveria et al., 1995). This increase in prevalence of OA is related to the ageing process and a reduction in neuromuscular joint protective mechanism, increased joint instability and a reduction in the resiliency of joint cartilage caused by a decreased anabolic response to growth factors (Arden & Nevitt, 2006). Ageing as a risk factor

also interacts with other risk factors to amplify the increasing risk of developing OA (Loeser, 2011).

3.4.2 Gender

Females have an increased risk of developing OA in the knee particularly later in life compared to men (Neogi & Zhang, 2013). In UK primary care females were found to have a higher incidence of consulting healthcare for their knee OA compared to males across all age groups (ARUK, 2013). Hormones may be responsible for such increased risk, particularly after the menopause when a deficiency in oestrogen may cause a higher risk to knee OA (Loeser, 2011). Although ageing and being post-menopausal in females increases the risk of developing OA, deficiency in oestrogen is not a clear risk factor in OA (Nevitt & Felson, 1996). There has been a mixed response to supplementation of oestrogen post-menopause with some radiographic images showing joints with a higher amount of cartilage with supplementation as compared to controls (Richette et al., 2003). There was no difference between the treatment and control groups in terms of joint pain symptoms (Arden & Nevitt, 2006). Although the prevalence of knee OA has been shown to be higher in females, the prevalence of OA in the hip and hand are similar in both genders across adult age groups (Pereira et al., 2011).

3.4.3 Obesity

Obesity has been suggested to be one of the leading risk factors in knee OA (Felson et al., 2000). Obesity as a risk factor for OA causes concern for public health in the UK, due to the increasing number of overweight and obese adults in the UK each year (Rennie & Jebb, 2005). In the UK it has been estimated that

more than 60% of adults have a body mass index (BMI) greater than 25.0kg.m^2, with this percentage set to increase to 70% by 2020 (Wang et al., 2011). An American case-control study of female adults indicated that having a high BMI increased the odds of having hand OA from 1.0 at normal BMI to between 6.8 and 9.3, hip OA to from 1.0 to between 1.4 and 3.4 and knee OA from 1.0 to between 3.8 and 9.3 (Oliveria et al., 1999). Obesity causes increased risk of OA due to the mechanical stress that increased weight causes on the joint, as well as metabolic factors affecting the structure of the joint (Felson & Chaisson, 1997). Obesity can exacerbate the development of OA leading to subsequent structural deterioration of the knee joint (Berenbaum et al., 2013). Obesity has also been shown to be a risk factor in the development of hip OA, although, the risk is lower compared to knee and hand OA (Oliveria et al., 1999). This is due to increased weight bearing that some joints take as part of obesity, creating excessive mechanical loading on cartilage and causing damage to ligaments in the knee, although the mechanism for the hand is not clear (Felson et al., 2000; Aspden, 2011).

3.4.4 Acute injury and joint malalignment

Injuries to joints including damage to the cruciate ligament, fractures and dislocations increase the risk of subsequently developing OA (Zhang & Jordan, 2010). In addition to injury causing OA; abnormal loading forces within a joint can cause damage or remodelling resulting in malalignment of the joint which may also increase the risk of OA if the remodelling is sub-optimal (Gelber et al., 2000). For example, hip impingement is strongly associated with developing hip OA in later life even after treatment of impingement (Ganz et al., 2003).

Continual high loading of joints in specific activities has been shown to be a risk factor for OA on the joints involved. Activities that require kneeling, squatting or heavy carrying can cause increased risk of OA in the knee and hip (Coggon et al., 2000). The hand is also at increased risk of OA in activities that require a pincer grip with the fingers and thumb (Arden & Nevitt, 2006).

Data from both the osteoarthritis initiative (OAI) study and the multicentre osteoarthritis (MOST) study which investigated if PA increases the risk of developing OA (Felson et al., 2013) found high mechanical loading and trauma to be risk factors for developing OA and both can occur in PA. Using the PASE questionnaire, cross-sectional data were analysed for the risk of developing radiographic knee OA in those who reported the highest levels of PA compared to those reporting lower levels of PA. Findings showed no differences in radiographic incidence of OA in the higher PA group compared to lower, suggesting that when excluding cases of previous knee injuries; PA causes no higher or lower risk of developing OA (Felson et al., 2013). Elite athletes have an increased risk of developing OA in joints that suffer high impact forces and increased joint loading independent of joint injury connected to their sporting activities (Buckwalter & Lane, 1997). Knee injuries in adults aged 55 and over have shown to increase the odds ratio of developing OA later compared to controls with no history of knee injuries (Conaghan, 2002). A Finnish study on the risk of OA in an international level sports athlete cohort (n=2000) found that long distance runners, endurance skiers, soccer, ice hockey and basketball players, track and field athletes, boxers, wrestlers and weightlifters were all at increased risk of requiring hospital care in relation to OA over a 21 year period compared to matched healthy control adults (Kujala et al., 1994).

Abnormalities within a structure of a joint can cause different distributions of forces acting upon it which in turn can lead to a higher risk of OA within the affected joint. Valgus joint deformity is the outwards angulation of the distal portion of a joint (Karachalios et al., 1994). In the knee, although a valgus joint is associated with increased risk of OA, varus joint deformity has a higher risk compared to valgus (Cerejo et al., 2002; Cahue et al., 2004). Varus joint deformity is the inwards angulation of the distal portion of a joint (Karachalios et al., 1994). An example of this is where the alignments within the knee joint dictate the effect of load distribution of the knee. If the knee is misaligned then this has shown to affect the joint cartilage and increase the risk of OA within the knee (Anderson et al., 2011). Malalignment can be caused by repetitive activities; this has been particularly observed in some occupations. In the hand, adults whose work required more pincer grip type actions had high risk of developing OA compared to those using the power grip (Hadler et al., 1978). The risk of knee OA is much higher in those with manual jobs that require regular squatting, kneeling or lifting of heavy weights (Coggon et al., 2000). Farming workers that have been in such an occupation for over 10 years are at a much higher risk of developing hip OA compared with adults with sedentary work (Croft et al., 1992).

3.4.5 Muscle strength and muscle weakness

In some joints, increased muscle strength has been shown to increase the risk of OA due to the increase in forces acting within the joint. An example of this is where an increase in hand grip strength in males increases the risk of developing hand OA (Arden & Nevitt, 2006). Muscle weakness is also associated with OA and is usually due to atrophy (McAlindon et al., 1993). Muscle weakness in the

quadriceps, even in the absence of muscle atrophy, is common in adults aged 45 years and over with knee OA with this association being stronger in females compared to males (Slemenda et al., 1998). A case controlled study comparing adults with OA in the hip, to age and gender matched healthy controls, found that the isometric strength and cross section of muscle mass in the thigh and around the acetabulum in those with OA was much lower compared to the control group (Arokoski et al., 2002)

3.4.6 Genetics

Genetic factors contribute to the susceptibility of developing OA in the foot, knee, hip and hand in females and the hip in males (Lanyon et al., 2000); risk of OA in different joints is predominately determined by genetic factors (Felson et al., 2000). In knee OA (Valdes et al., 2008), 17 different genotypes were identified for risk of OA in a case control study. Genetic factors are a stronger determinant for OA compared to that of obesity and previous injury (Valdes et al., 2008). Many genes play a role in the development of OA, some of which have been found to effect growth of type II collagen which forms the joint's cartilage thus making it susceptible to developing OA (Arden & Nevitt, 2006).

While some of these risk factors have been identified as independent for the development of OA, interactions between more than one of these risk factors are the most likely cause of OA in individuals, as they may not be interdependent (Blagojevic et al., 2010). For example, injury would increase risk of OA alone, an injury would lead to risk to the joint becoming malaligned and reduced muscular strength, therefore, risk factors should not be thought of as individual factors that only influence the chances of developing OA (Neogi & Zhang, 2013).

3.5 Pathology of osteoarthritis

OA can be defined pathologically, radiographically or clinically, by either symptoms within a joint or by its structural pathology, although it is most useful to use a combination of the two approaches (Altman et al., 1986; Felson et al., 2000; Pereira et al., 2011; NICE, 2014). The main symptoms reported by the individual with OA include joint pain, stiffness or reduced mobility of the joint, causing physical disability to the individual (Dieppe & Lohmander 2005). The pathophysiology that results in the symptoms of OA can vary to include loss of joint cartilage and joint space narrowing, osteophyte growth at the joint, weakening of muscles supporting the joint and occasionally joint inflammatory synovitis leading to remodelling of the joint (Pelletier et al., 2001).

The pathological definition of OA is defined by changes to the tissue that begin with alteration of the joint's ability to repair damage to the cartilage caused by mechanical forces acting on the joint (Intema et al., 2010). The failure to repair the damage is caused by biomechanical and biochemical changes in the joint that prevent or limit nutrients and oxygen reaching the extracellular matrix of the cartilage (Bijlsma et al., 2011). This leads to a cartilage matrix that is unable to withstand normal mechanical stress which causes further damage to the cartilage matrix leading to the unsuccessful repairing of the joint (Bijlsma et al., 2011).

In addition to cartilage damage, the role of the bone and synovial tissue around the joint have become evident in the development of OA. Pathological changes that occur within the bone in OA are characterised by osteophyte formation at the subchondral bone, subchondral bone growth, sclerosis and attrition of the subchrondral bone as a response to mechanical stresses to the joint (Hayami et

al., 2006). Attrition of the subcondral bone can lead to abnormal shaping of the articular bone such as flattening or depression (Neogi & Zhang, 2013). Bone marrow lesions have also been noted in studies caused by increased mechanical forces acting on bone (Hunter et al., 2008). Bone marrow lesions are where there have been high amounts of cell necrosis, fibrosis and remodelling where the bone is repairing (Felson et al., 2001).

Synovial inflammation or synovitis occurs in OA and relates to the clinical symptoms of OA which includes joint swelling and pain due to inflammation (Bonnet & Walsh, 2005). Synovitis is thought to be caused secondarily to other changes in the joint as cartilage is damaged and other inflammatory mediators affecting the synovial cavity (Bijlsma et al., 2011). This inflammation can further cause problems in the cartilage matrix, affecting the damage in the cartilage as the synovial tissue inflammation pushes the cartilage and also affects the repair response of the cartilage; although cartilage, joint repair and remodelling can occur, inflammation of synovial tissue can lead to disease development within the joint (Sellam & Berenbaum, 2010; Bijlsma et al., 2011). By damaging the cartilage, more significant inflammation can occur in synovial tissue leading to a worsening cycle of joint damage and symptoms, the inflammation in OA is not as severe as in rheumatoid arthritis (Bijlsma et al., 2011).

This pathological model suggests that OA is due to unusual, high mechanical loading of the joint leading to inflammation, joint and bone damage. There has been an alternative hypothesis that suggests inflammation has a much higher role in the development of OA (Bondeson et al., 2006). There is evidence that biomarkers of inflammation are produced within the subcondral bone, the cartilage

and synovial tissue in early OA and later stages of development of OA (Wang et al., 2011). This alternative hypothesis states that alongside mechanical factors causing the development of OA, there are also systematic factors with inflammatory biomarkers being produced within and around the joint. This causes the inflammation and pain in the synovial tissue or, other tissues in the joint (Pelletier et al., 2001). Studies have supported this argument at a genetic level in animal models with genes associated with better anti-inflammatory responses having less cartilage loss and bone damage when OA is induced by a mechanical force (Van der Kraan., 2012). Another group of tissue affected by OA around the joint is that of the skeletal muscle that is associated with the joint and the ligament (Lohmander et al., 2007). In OA, muscles have been found to have become weaker in strength with lower cross sectional size and density, containing larger number of fatty deposits (Bennell et al., 2008). Muscle weakness is thought to be caused by a combination of the natural development of OA and the decrease of use due to the joint pain symptoms of OA (Loureiro et al., 2013).

3.6 Diagnosis of osteoarthritis

3.6.1 Radiographic

Changes within the joint caused by OA can sometimes be seen using imaging techniques (Dieppe et al., 2000). The American College of Rheumatology (ACR) definition for diagnosis of radiographic OA is for the x-ray to display damage to cartilage and narrowing of the gap between the bones as well as osteophyte formation that can be observed in the subcondral bone (Altman et al., 1986; Altman et al., 1990; Altman et al., 1991). The limitation of a radiographic diagnosis

of OA is that there are many cases of symptomatic OA with no sign of radiographic evidence for joint changes; this may link with inflammation as pathology of OA or the experience of joint pain even when structural damage is not present (Dieppe et al., 1997). There are also cases where radiographic changes have occurred in the joint but symptoms are stable without pain or stiffness (Peat et al., 2001). Due to this limitation, NICE do not recommend using radiographic investigation of patients that meet the clinical diagnosis of OA in primary care (NICE, 2014).

3.6.2 Clinical diagnosis

In view of the issues with the radiographic definition, it may be best to consider the clinical definition of OA rather than the physiology behind it and focus on the symptoms associated with OA (Peat et al., 2001).

A clinical definition of OA is based on the symptoms the individual is experiencing and the risk factors for developing OA (Lawrence et al., 2008). The National Institute for Health and Care Excellence (NICE) recommend clinicians use a working diagnosis of peripheral joint OA in their most recent guidelines (NICE, 2014). A working diagnosis is usually the preliminary diagnosis based on symptoms and physical signs that can be made to allow for early management of OA (NICE, 2014). Box 3.1 displays the criteria for a working diagnosis of OA recommended by NICE (2014).

Box 3.1 Criteria for NICE working diagnosis of OA

Individuals reported joint pain that was persistent and became worse with use.
Aged 45 years and over
Individuals report no morning joint-related stiffness, or morning stiffness that lasts no longer than 30 minutes

This definition of clinical OA is similar to that recommended by the ACR which differentiates OA from other inflammatory definitions (Altman et al., 1986); the ACR clinical definition also includes x-ray. The ACR clinical definition for knee OA includes examining for crepitus, morning stiffness, bony enlargement and age at 38 years and over (Altman et al., 1986).

3.7 Symptoms of osteoarthritis

Persistent and long term joint pain, short episodes of stiffness in the joint and overall reduced joint function are symptoms commonly experienced by individuals with OA (Felson et al., 2000; Neogi & Zhang, 2013). Joint pain symptoms are usually localised in the affected joint (de Bock et al., 1995). Pain is often reported relating to joint use, although resting joint pain and night time joint pain associated with OA has also been observed in some individuals; the pain experienced by patients with OA in the affected joint includes tenderness, aching or throbbing. Individuals have stated that joint activity induced pain can result in episodes of sharp and stabbing pain (de Bock et al., 1995).

More advanced knee OA has also been associated with higher joint stiffness (Zeni & Higginson, 2009). In adults diagnosed with severe knee OA, severe joint stiffness has been observed, with individuals experiencing significant knee restriction in flexing and responding to weight loading (Ramsey et al., 2007). Severe OA and higher amounts of joint stiffness have been found to be associated with lower self-selected speeds of walking, longer stride duration, shorter stride length and less steps per minute compared to healthy controls (Zeni & Higginson, 2009).

Limited range of motion in the knee and hip joint has also been found to be correlated with increased disability in those with OA (Steultjens et al., 2000). It is reported that this limited range in motion accounts for 20-25% of the variance in disability among OA populations (Steultjens et al., 2000).

Pain and stiffness in the joint which increases in severity as OA progresses can lead to high levels of physical disability, affecting the physical and mental health status of individuals suffering from OA. Physical disability in individuals with OA has been seen to affect individuals' physical and mental health-related quality of life (Salaffi et al., 2005). In those with knee OA, high pain scores have been associated with increased risk of depression, anxiety, stress and poorer mental health scores, showing the severe effect pain in OA can have on individuals (Creamer & Hochberg, 1998). Those with OA are more likely to report poor physical functioning compared to those with no OA (Ettinger et al., 1994). In those with knee OA, 50.0% males and 70.5% females reported difficulties walking on a level gradient or walking up or down at least two steps, compared to 18.8% males and 24.8% of females with no OA (Ettinger et al., 1994). In reporting difficulties transferring from sitting positions or from bathtubs, 43.8% males and 67.2% females with OA reported difficulty compared to 18.5% males and 27.5% females without OA. In adults aged 41 years and over with reported OA symptoms in the knee or hip, 27% stated they do not walk up or down hills, 26% walked shorter distances or stopped to rest often due to joint pain with 32% going up and down stairs more slowly (often one step at a time) (de Bock et al., 1995).

3.8 Management of osteoarthritis

OA is an irreversible progressive health condition; treatment and health management for OA is not aimed at repairing joints but to minimise disability and symptoms as well as reducing OA progression. The NICE guideline for the treatment and management of OA recommendes: education and self-management, non-pharmacological management, pharmacological management, considerations for surgery and patient follow up. The essence of the NICE guidelines for OA management is a holistic approach to patients, allowing for assessment of the individual's medical, social and psychological needs and building a self-management intervention from the recommended treatments relevant to those needs (NICE, 2014).

3.8.1 Education and self-management

NICE recommend all people with OA are offered precise and complete information about OA to enable their own understanding of the condition and its management. Information and education together with this management can include written and oral information to counter misconceptions, progression and treatment options (NICE, 2014). Education and information sharing should be a constant process rather than at a singular time point (NICE, 2014). The self-management should be designed to encourage patients to be active in promoting their own health, examples of this include exercise for OA which can improve patients' outcomes. This principle encourages patients to have the knowledge and skills to self-manage.

Resting and pacing appear to be beneficial when a joint hurts, this maybe the case in acute pain due to injury, muscle atrophy is a feature of OA and so exercise is recommended rather than resting. Using thermotherapy, either hot or cold is recommended as an adjunct to core treatments as a part of self-management if patents gain benefit from it (NICE, 2014).

3.8.2 Non-pharmacological management

The non-pharmacological management of OA recommended by NICE includes exercising, weight loss (where necessary), footwear or cushioning in soles for shoes. The NICE guidelines also recommend that assistive devices and transcutaneous electrical nerve stimulation (TENS) electrotherapy could be used complementarily with the core treatments of education and self-management, exercise and weight loss. All people with OA should participate in exercise as part of their self-management of OA regardless of age, other co-existing health conditions, pain or level of disability (NICE, 2014). Exercises recommended are local and muscle strengthening together with general aerobic fitness activities. The NICE guidelines do not make recommendations on how exercise interventions should be provided to patients. In overweight or obese patients, weight management should be offered as excess weight loading on OA joints are a major factor for progression (NICE, 2014). There is some evidence for TENS in short term reduction of pain, NICE recommend that TENS treatment can be offered alongside the core treatment recommended for all patients (Rutjes et al., 2009; NICE 2014). As part of core treatment for those with lower limb OA, advice on correct shock-absorbing footwear should be given and as adjunct to core

treatment assessments for assistive devices such as walking stick, braces or joint support, if appropriate, depending on the patient's preferences.

3.8.3 Pharmacological management

In addition to the core treatments for OA, in 2008 NICE (NICE, 2008) also recommend healthcare professionals consider offering paracetamol for pain relief. A regular dose of paracetamol along with topical non-steroidal anti-inflammatory drugs (NSAIDs) should be considered before oral NSAIDs, cyclo-oxygenase-2 (COX-2) inhibitors or opioids. Other opioid analgesics should be only considered if topical NSAIDs and paracetamol are not effective for patients, although the risk of administration should be considered (NICE, 2014).

3.8.4 Considerations for surgery

Joint replacement surgery should be deliberated if patients have already received core treatments for OA and their joint symptoms are having an impact on their quality of life even with the core recommended treatments. Joint replacement surgery consideration should be prior to joints having progressed to functional limitation and severe pain; recovery period and rehabilitation for patients should be evaluated prior to surgery (NICE, 2014).

3.8.5 Patient follow-up

The NICE guidelines recommend that all patients with OA are offered a regular review of their symptoms. The timing of the follow-up review depends upon the individual; this should be given for those that have joint pain that is interfering with their daily living, more than one joint with joint pain symptoms, more than one

comorbidity or are taking regular pharmacological treatment for their OA (NICE, 2014).

In 2014, NICE guidelines highlighted the importance of PA and exercise as a core treatment in the management of OA for all adults aged 45 years and over with the condition. PA and exercise were highlighted as important in the management of OA as it can manage the symptoms of joint pain and loss of function (NICE, 2014).

3.9 The benefit of physical activity and exercise in the management of osteoarthritis symptoms

OA is one of the most common causes of physical disability among adults aged 45 years and over in Europe and the United States (Woolf & Pfleger, 2003). Exercise is effective in reducing pain and increasing physical functioning in adults aged 45 years and over with OA in the lower limbs (Uthman et al., 2013). Exercise interventions for hand OA have shown to be limited in their effectiveness for reducing hand pain (Dziedzic et al., 2011; Kjeken et al., 2011). In the most recent NICE guidelines for the management of OA of the knee, hip, foot or hand, strengthening exercises and general PA are recommended regardless of age, disability or co-morbidity (NICE, 2014). PA and exercise has also been recommended as treatment for OA symptoms internationally by the American College of Rheumatology (ACR), European League Against Rheumatism (EULAR) and OA Research Society International (OARSI) (Hochberg et al., 2012; Fernandes et al., 2013; McAlindon et al., 2013). A Cochrane meta-analysis of exercise intervention for adults with knee OA reported that exercising can have a similarly strong effect at reducing pain symptoms to that of analgesics (Fransen &

McConnell, 2008). NICE updated its 2008 guidelines for OA and aerobic PA and muscle strengthening exercises were recommended as a core treatment for all adults with OA irrespective of age, comorbidity, pain severity or disability in both guidelines (NICE, 2008 & 2014). PA and exercise is also recommended by the European League Against Rheumatism (EULAR), the Osteoarthritis Research Society International (OARSI) and the American College of Rheumatology (ACR) in their guidelines for the management of OA (Hochberg et al., 2012; Fernandes et al., 2013; McAlindon et al., 2013).

There is large amount of evidence for the positive benefits of PA in terms of reducing pain symptoms and improving function (Bartels et al., 2007; Fransen & McConnell, 2008; Fransen et al., 2014). Trials on PA interventions for individuals with knee OA were published from the early 90s (Kovar et al., 1992; Börjesson et al., 1996; Bautch et al., 1997; Ettinger et al., 1997; O'Reilly et al., 1999). Kovar et al. (1992) conducted a trial of 102 patients with OA in one or both knees comparing an 8-week supervised walking intervention compared to a non-exercise usual medical care control. The walking intervention group demonstrated a decrease in OA joint pain and an increase in physical function in walking distance compared to the control group (Kovar et al., 1992). A large study of 786 participants in a four-arm RCT was conducted in the UK to examine the long-term benefits and adherence of home-based resistance exercises for knee OA (Thomas et al., 2002). At the 24 month follow up 600 participants completed the study with resistance training showing to be highly significant in reducing knee pain compared to non-exercise controls, with greater adherence to exercise shown to lead to more effective pain relief (Thomas et al., 2002). A more recent RCT investigated an integrated PA intervention against usual care control (Hurley et al.,

2012). Even after the 30 month follow up, the intervention arm still had significantly higher physical function and higher cost benefit in health and social costs (Hurley et al., 2012).

Other RCTs then emerged to compare different types of physical activities. These included comparing the effect of different forms of activities or evaluating the effectiveness of combinations of activities in reducing OA symptoms and increasing physical functioning (Ettinger et al., 1997; Fransen et al., 2007; Chaipinyo & Karoonsupcharoen, 2009). Ettinger et al. (1997) conducted a three-arm RCT comparing the effectiveness of aerobic walking, with resistance training and health education in 365 adults with knee OA. Both the aerobic walking and resistance training interventions were more effective compared to the health education intervention at reducing self-report pain symptoms, disability and improving performance on physical functioning tests (Ettinger et al., 1997). In this trial, no significant difference in resistance strengthening exercises and aerobic PA was found in their effectiveness for treating knee OA.

Fransen & McConnell (2008) conducted a systematic review and meta-analysis of all exercise intervention trials for knee OA and concluded that the effectiveness of exercise interventions had similar effect sizes in reducing pain and disability to that of pharmacological analgesics and non-steroidal anti-inflammatory drugs; this comparison of benefit is only specific to OA symptoms. When considering the wider health benefits of PA and/or exercise compared with the safety and possible negative side-effects of pharmacological drugs, PA would seem highly favourable (Powell et al., 2011; NICE, 2014).

It is important to note that there is a common misconception about exercise and PA in OA (Roddy et al., 2005). This is that exercise will further damage and 'wear out' joints with OA due to the pain patients may experience during exercise. Evidence shows that for those who continue with exercise programmes, the vast majority do not suffer any worsening of symptoms even in those with severe symptoms (Roddy et al., 2005).

The benefits of PA have been shown in both strengthening exercise on localised joints as well as general aerobic exercise in OA patients for improved change in pain symptoms and decreasing physical disability (Fransen & McConnell, 2008).

A recent meta-analysis illustrated that there is sufficient evidence to demonstrate that PA interventions offer clinical benefit in lower limb OA for pain symptoms and physical functioning, with the majority of this evidence existing for knee OA (Uthman et al., 2013). The meta-analysis suggested that strengthening exercises or combinations of different forms of PA are the most effective in lower limb OA pain and disability (Uthman et al., 2013). The meta-analysis on the effectiveness of different forms of activities showed that aqua-based strengthening and aerobic activities or strengthening and flexibility were most effective for reducing pain and increasing function (Uthman et al., 2013). Combined interventions of land-based strengthening, aerobic and flexibility activities were more effective than interventions of single forms of activities and of which strengthening activities seemed to be most effective. Flexibility activities were less efficacious in both increasing function and reducing pain symptoms compared to strengthening. Aerobic activities interventions showed as less beneficial in reducing pain

symptoms as flexibility and strengthening activities but were similar in increasing function (Uthman et al., 2013).

The network meta-analysis indicated that aqua-based combinations of different forms of PA compared equally to strengthening exercises or land-based combination exercises. A RCT conducted in the UK of n=312 patients with knee or hip OA, showed that aqua-based exercise interventions; set in public swimming pools, to be more efficient in pain and physical function compared to a usual care control (Cochrane et al., 2005). The exercise intervention included strengthening and range of motion flexibility exercises along with cardiovascular aerobic activities during one hour sessions, twice a week. The net reduction in cost saving per patient per year of the intervention was estimated at £123-175, and an incremental cost-effectiveness saving of £3838-5951 per quality adjusted life year (Cochrane et al., 2005).

The mechanism for the benefits of PA in OA and joint pain is thought to be by improving the structure and tissues around the joint, as a physiological response to PA (Øiestad et al., 2013). This strengthening of structure and tissue around the joint, including skeletal muscle, increased blood supply and neuromuscular control in its ability to reduce mechanical strain associated with joint pain in OA (Iwamoto et al., 2011).

3.10 Summary

This chapter has examined the burden of OA in the UK and the threats to public health, the importance of PA in the treatment of OA symptoms have also been discussed. The main features of this chapter were:

- OA is highly prevalent in the UK and is the fifth leading cause of disability in the UK
- OA is a clinical syndrome characterised by joint pain
- PA is a core treatment for the management of OA in reducing joint pain and disability
- Measuring levels of PA in adults aged 45 years and over with OA would determine if further PA interventions are needed

Chapter four: Measuring levels of physical activity

4.1 Introduction

Chapter three described the importance and usefulness of measuring levels of PA in adults aged 45 years and over with joint pain and OA; one challenge of such measurement is selecting the appropriate approach. This chapter will address:

- The different approaches of measuring PA
- Selecting self-report measures for assessing PA
- Describing classical test theory
- Analytical techniques for assessing measurement properties
- Measurement properties of self-report PA instruments

4.2 Measuring physical activity

PA is recommended by various guidelines across the world. It is an important indicator of health status and has been shown to lower pain and disability in adults with joint pain or OA (Sun et al., 2013; Uthman et al., 2013). Measuring PA in adults aged 45 years and over with joint pain is important in describing levels of population participation in PA; assessing how many adults aged 45 years and over with joint pain are physically active. If adults aged 45 years and over with joint pain are participating in PA, measuring the amount can identify possible factors associated with its reduced levels, so that groups that are at risk of being inactive can be identified and interventions can be tailored (Stubbs et al., 2014). Measuring

PA is also important for tracking changes in levels of PA over time (Svege et al., 2012). Quality measurement of levels of PA in this population allows for the reliable estimates to be established and this can also be used to monitor trends in PA levels over time (Montoye, 2000) and study relationships between PA and health outcomes (Terwee et al., 2011).

4.3 Different approaches for measuring physical activity

There are a variety of approaches to measuring PA, each with different properties of measurement, strengths and limitations. Different approaches are appropriate to different settings (Ainsworth, 2009). Each approach also varies in its level of precision; methods with high precision are normally less practical for measurements in daily living (Lamonte & Ainsworth, 2001). Table 4.1 displays the seven different methods of measuring PA used within research and an overview of their precision of measurement and ease of practical application (Ainsworth, 2009).

Table 4.1 Different methods for measuring levels of PA in research.

	type of method	Level of precision of measurement	Ease of practical application
Objective Measures	Room Calorimetry	High ↓ Low	Low ↓ High
	Double Labelled Water		
	Indirect Calorimetry		
	Heart Rate Monitoring		
	Accelerometry		
	Pedometers		
Subjective Measures	Self-Report Questionnaires		

4.3.1 Objective measures

Objective measures of PA do not rely on subjective information provided by the individual; instead measurement and recording of biomechanical, physiological responses to performing PA are made, often in real time (Trost & O'Neil, 2014). Objective measures are not at risk to reporting bias or recall bias that is often associated with subjective measures (De Vet et al., 2011). The precision of different objective measures depends on the biomechanical, physiological responses that are measured and recorded (Lamonte & Ainsworth, 2001).

(1) Room calorimetry

The most precise instrument for measuring PA is room calorimetry or direct calorimetry, where the study participant stays in a room and all of the gas in the room is analysed in terms of oxygen and carbon dioxide levels (Seale & Rumpler, 1997). Measurement of oxygen and carbon dioxide within the room can be used to estimate the amount of PA the subject participates in during the time spent in the room (de Jonge et al., 2001). The advantage of room calorimetry is the precision in its measurement of PA and is currently the best available. Another advantage of room calorimetry is the measurement unit of PA is in energy expenditure over time, making measurements interpretable to time spent active over a day. One key limitation of room calorimetry is that it can only assess one subject at a time. The equipment is expensive requiring trained technicians to collect data and is restricted to activities that can be conducted within the room and therefore is not a reflection of free living PA (de Jonge et al., 2001). Despite this method's high accuracy, it is limited in its ease of use and therefore this method would be best

suited to laboratory research in PA rather than gaining understanding of PA at a population level or within clinical trials.

(2) Double labelled water

Double Labelled Water (DLW) is another example of a method for measuring PA that is highly accurate but expensive. DLW method uses specialised water which has had the hydrogen and oxygen particles replaced with non-radioactive marked hydrogen and oxygen particles, the subject consumes the specialised water, these particles are absorbed as metabolism takes place when the subject is physically active (Westerterp, 2009). Energy expenditure can be estimated by the concentration the non-radioactive marked particles, which can be measured by sampling saliva, urine or blood (Westerterp, 2009). The advantage of the DLW method is that energy expenditure can be measured for a set time period (e.g. days to weeks) and offers an accurate measurement of PA as energy expenditure in free-living rather than with restrictions like room calorimetry (Westerterp, 2009).

DLW is not only an expensive method to use (Plasqui & Westerterp, 2007) but also total energy expenditure can only be collected from saliva, urine or blood samples at the end of the measurement period. Data cannot show timelines of intensities of PA, it can only be collected as total energy expenditure between time periods (Plasqui & Westerterp, 2007).

The DLW method is viable for collecting energy expenditure data in free-living for a single study participant (Ainsworth, 2009), not practicable for a large population study due to costs and an inability to measure isolated PA patterns over a set time period; this limits its practical use in population based research trials.

(3) Indirect calorimetry

Indirect calorimetry is a less precise method but is less expensive for measuring PA compared with room calorimetry and DLW (Livesey & Elia, 1988). Indirect calorimetry is the measurement of oxygen and carbon dioxide as a proxy measurement of energy expenditure (Livesey & Ella, 1988). Gases expired are measured using open or closed circuit spirometry from which computer software can use algorithms to estimate energy expenditure (Livesey & Elia, 1988). Indirect calorimetry has been shown to be moderately accurate in estimating energy expenditure with small variance in measurement compared to room calorimetry and DLW (Levine et al., 2000). Indirect calorimetry is limited due to the required spirometry equipment which is immobile and restricted to a laboratory setting (Levine et al., 2000). Whilst there have been advances in technology allowing for mobile spirometry, a mouthpiece is needed to be worn at all times during measurement and would not be practical for measuring daily PA (Bonomi et al., 2010). Indirect calorimetry is best used for individual measurements of exercise in a laboratory setting.

(4) Heart rate monitoring

A linear relationship between heart rate and energy expenditure has been demonstrated during constant workloads that use large muscle groups (Strath et al., 2000), for example steady speed cycling (Ceesay et al., 1989). Heart rate monitoring can also be used in research to monitor daily PA of adults in free-living (Achten & Jeukendrup, 2003).

Estimating PA using heart rate monitoring allows for a less invasive method compared to indirect calorimetry, DLW and room calorimetry. When individually calibrated, heart rate monitored estimations of energy expenditure have been shown to be acceptable and accurate compared to DLW in free-living adults (Livingstone et al., 1990).

The limitations of using heart rate monitoring are that it can be time consuming and costly to individually calibrate heart monitors for each subject in larger scale studies (Spurr et al., 1988). Inaccuracies can occur in heart rate monitoring because there are other factors that cause changes in heart rate separate to increasing PA, for example, stress, cardiovascular drift (changes in heart rate during same activity workload), hydration, and rising and falling body temperature (Achten & Jeukendrup, 2003).

Heart rate monitoring has demonstrated inaccuracies in measuring PA compared to DLW, nevertheless it is considered an acceptable approach to accurately measure PA (Trost & O'Neil, 2014). Devices for measuring PA using heart rate monitoring in large populations are costly making them currently impractical, advances in technology are making heart rate monitoring a more feasible approach with reduced costs for the future (Trost & O'Neil, 2014). Measuring PA with heart rate monitoring is currently too costly and impractical to implement at a population level, nonetheless it may become a viable option in the future with technological advances making Heart Rate Monitoring more accessible and cheap (Strath et al., 2013).

(5) Accelerometry

Accelerometry measures direct movement of the body via a device attached to specific areas of the body. It measures the movements and estimates the muscle workload required for the movement over distance and speed, resulting from this energy expenditure can be estimated (Chen & Bassett, 2005). A variety of accelerometer devices are available and can measure movement in one, two or three axes depending on the different brands and models (King et al., 2004). Unlike heart rate monitoring, accelerometry doesn't require calibrations for each individual subject. Although accelerometers' usability varies in different models, they are generally easy to use.

Accelerometry can provide details of PA in free-living conditions and can display timelines and patterns of PA at different intensities and periods of rest (Chen & Bassett, 2005). Accelerometers only measure the movement of the body and therefore would not give details of carrying loads, cycling exercise and walking up and down gradients (Johannsen et al., 2010). Using accelerometry in research can be costly with the more sophisticated the device the higher the cost. Obtaining complete data of accelerometers can be an issue in research. In postal surveys, activity monitors need to be sent through the post and rely on the subjects to return them similarly and the postal service to return them all safely and undamaged. For example, an American survey distributed 7176 adults accelerometers but only 4867 (67.8%) returned the accelerometer with four days of data (Troiano et al., 2008). This makes using accelerometers currently an expensive option, with a risk of low response rates.

(6) Pedometers

Pedometers do not measure levels of PA but measure the number of steps taken in a given time period and have shown to vary considerably in their test-retest reliability and validity (Colbert et al., 2011). Some pedometers have been shown to underestimate the number of steps in slower and short periods of walking compared to observed measurements of walking (Terwee et al., 2011). Pedometers are limited in their ability to distinguish different intensities of PA, which is a key measure for health outcomes (Corder et al., 2007) and that the measurement is only given in steps. There have been studies that have made recommendations in the number of steps per day for adults (Haskell et al., 2007; Ewald et al., 2013). The recommended numbers of steps are restricted, as evidence shows that the number of steps that equates to the same amount of PA varies in different populations depending on age, body composition and physical fitness (Ewald et al., 2013). Pedometers may be considered an appropriate measure for levels of PA in populations, as pedometers are relatively cheap to buy, user friendly and are objective with a quantifiable outcome score. Although cheap and easy to administer for objectively measuring activities, pedometers are not a reliable or valid method for measuring levels of PA due to their limitations. Pedometers have been successfully implemented as a tool for increasing exercise adherence, where individuals track their daily number of steps as part of a graded PA programme, giving pedometers an important role in terms of adherence to PA rather than as a measurement instrument (Stovitz et al., 2005).

4.3.2 Subjective measures

Subjective measures of PA include patient reported outcome measures, such as self-report questionnaires. These self-report instruments require individuals to recall and report their activity retrospectively (De Vet et al., 2011). Subjective measures have been used to measure other parameters such as health status (Gandek et al., 1998), joint pain-intensity (Zelman et al., 2005) and assessing risk of mental health problems (Zigmond & Sniath, 1983) (Mokkink et al., 2010). Subjective measures are commonly used to measure unobservable phenomenon (De Vet et al., 2011). Given PA is observable but requires observations over long periods of time, subjective measures are preferable. Subjective measures of PA are at risk to recall bias, reporting bias and social-desirability bias, whereas objective measure avoid these biases (De Vet et al., 2011).

(1) Self-report questionnaires

Self-report PA questionnaires are a popular approach for measuring levels of PA in larger population studies (Helmerhorst et al., 2012). This is due to their ease of use and immediate access to information about an individual's PA and the low cost involved in administering to study participants (Warren et al., 2010). There are many different types of PA questionnaires each with a specific aim and particular use with different groups, ages or in different contexts (van Poppel et al., 2010). An example of a questionnaire with a specific purpose is the PA Scale for the Elderly (PASE) which was developed specifically for measuring levels of PA in adults aged 65 years and over (Washburn et al., 1993). The selection of the correct PA questionnaire depends on the population being assessed, the evidence

of the measurement properties for that population, how the questionnaire will be administered and the interpretation for the results (Terwee et al., 2010).

A limitation with measuring PA using questionnaires is the precision of measurement in the questionnaires themselves. There are many studies which examine the precision in a variety of PA questionnaires in different populations; there is a lack of high quality studies on their accuracy and validity (Hart et al., 2011). This is particularly the case in the validation of the questionnaires in comparison to more direct measures like accelerometers, DLW, and calorimetry (Prince et al., 2008; Helmerhorst et al., 2012). Self-report instruments of PA would be useful for measuring population level PA, as they are cheap and quick to administer. There are wide selections of self-report instruments available depending of the age of the study participants, setting of PA or purpose of measuring PA. To reduce the risk of reporting and recall bias the appropriate self-report instrument must be selected according to the demographics of the participants (De Vet et al., 2011). Table 4.2 summarises the key characteristic of each of the seven methods explored. A common argument against self-report measure is that they are consistently not as good as objective methods (de Vet et al., 2011). In larger population studies, objective measures are not practical for measurement (de Vet et al., 2011). Self-report instruments are relatively inexpensive compared to more direct measures of PA and can be self-administered as part of a survey (Hart et al., 2011). These qualities make a questionnaire the most viable method for assessing PA in a large population, particularly as they can be used in postal surveys. When selecting a self-report instrument, the measurement properties using an established theoretical

framework should be considered to ensure best possible precision in measurement (de Vet et al., 2011).

(2) Classical test theory

When evaluating the measurement properties of self-reporting PA instruments, the classical test theory (CTT) approach was taken. CTT is one of the most widely used concepts in the development and evaluation of outcome measures in health research (McDonald, 1999). CTT is an approach that can be applied to measure observable and non-observable constructs. Examples of observable constructs include blood pressure, lung capacity and hand grip strength. Examples of non-observable constructs include health-related quality of life, mental wellbeing and self-efficacy (Streiner & Norman, 2008).

CTT is based on the assumption that in every measurement instrument there is an element of measurement error. This error could either be zero, or greater than zero. In the latter case, the test or measurement does not represent the 'true score'. The aim is to quantify the error, and determine its source. These sources can be defined as measurement properties.

While an individual's score in a measurement instrument is observable, the true score and measurement error is not. It is the aim of CTT is to quantify the measurement error within an instrument, in order to understand how closely an instrument's measurement reflects the individual's 'true' score of the construct (De Vellis, 2006). An important assumption of CTT is that each item is measuring the defined construct that it aims to measure, and that the measurement error in each

item is random and independent of all other items within the instrument (Streiner & Norman, 2008).

If an instrument is used to measure an individual in a given construct, the score of the instrument can be given as Y. The individual's 'true' score within the construct can be termed X, and the measurement error of the instrument can be given as E. The CTT approach then reflects this model as: Y=X+E (Streiner & Norman, 2008).

Quantifying the measurement error in an instrument is conducted by evaluating its measurement properties (De Vet et al., 2011). Previously, the definitions of different measurement properties varied within the literature. For example, some referred to reliability, whilst others used the terms reproducibility or repeatability (De Vet et al., 2006). In 2010, an internationally recognised set of definitions for measurement properties was published for health-related research, known as the COnsensus-based Standards for the selection of health status Measurement Instruments (COSMIN) (Mokkink et al., 2010). A summary of each of the measurement properties, and the techniques for assessing these properties, is given after this chapter (Section 4.5).

Table 4.2 Characteristics of methods for measuring PA

Method	Description	Unit of measurement	Strength	Limitation
Room Calorimetry	Monitors inspired and expired gases within a sealed room	Energy expenditure	Highest precision for measurement	Cost; restricted to one subject at a time; restricted to one room
Double Labelled Water	Measures energy expenditure based on uptake of marked molecules	Energy expenditure	Highly precise for measuring energy in a time period; can be used in normal living conditions	Cost; energy expenditure only available as a total count
Indirect Calorimetry	Estimates energy expenditure based on expired gases	Energy expenditure	High precision; monitors energy expenditure with data given immediately during activities	Requires constant mouthpiece and analysis equipment to be attached; expensive
Heart Rate Monitoring	Using heart rate monitoring to estimate energy expenditure	Energy expenditure	a precise and objective method for estimating energy expenditure over time; heart monitors are relatively affordable	Other factors can affect heart rate which can cause problems in precision; more sophisticated heart monitors can be expensive
Accelerometry	Uses accelerometers to estimate movement;	Depends on device; energy expenditure; activity count; steps	Objective method for measuring activities in daily life and can measure for up to 9 days depending on batteries	Vary in cost; validity varies in different devices
Pedometers	Accelerometers to measure number of steps	Step count	Objectively measure number of steps taken; cheap; can measure over many days	Questionable validity and reliability; number of step not valid in different populations
Self-Report Questionnaires	Patient Reported Outcome Measure	Depend on instrument used; energy expenditure; activity count; steps; time inactive	Cheap; easy to use and score immediately; used in a variety of setting; many available	Subjective; level of precision; measurement properties questionable

4.4 Outcome Measures in Rheumatology (OMERACT)

The OMERACT group is an international organised network with the aim of standardising outcome measures in RCTs and longitudinal observation studies in rheumatology (Boers et al., 2005). The OMERACT group recommends standardised outcome measures to be used in rheumatologically RCTs and longitudinal observation studies based on data on outcome measures and expert consensus. OMERACT together with OARSI created a set of outcome measures for clinical trials in OA. This included a measure of pain intensity, physical function and global perceived change in joint pain problems (Bellamy et al., 1997), from this an OMERACT-OARSI responder criteria was also developed (Pham et al., 2004). Currently an outcome measure of PA is not included in any of OMERACT recommendations (Bellamy et al., 1997). PA is an important outcome measure in adults aged 45 years and over with OA as it has been recommended as a core treatment for OA (Fernandes et al., 2013; McAlindon et al., 2014; NICE, 2014). This thesis is focussed on measuring and assessing PA in adults aged 45 years and over with joint pain. Findings are relevant to the OMERACT group in the assessment of PA as an outcome measure in OA clinical trials and observational studies.

4.5 Techniques for assessing measurement properties

When making choices for the most appropriate instrument, that instrument's measurement properties should be considered along with the instrument's qualitative attributes, which can include its clinical relevance. Evidence of measurement properties supplies the quantitative data on the questionnaires;

these measurement properties are categoried into groups of definintions (Mokkink et al., 2010).

Research in the area of measurement properties can become confusing due to a large variety of terminology that is used for specific measurement properties and how they are evaluated (de Vet et al., 2011). For example in 2003, a study identified many different methodologies in assessing responsiveness in health-related questionnaires that lead to different conclusions about responsiveness (Terwee et al., 2003). COSMIN agreed upon a uniform set of terms of measurement properties and study designs for assessing self-report instruments in the relevant properties (Mokkink et al., 2010).

(1a) Reliability

Reliability is the degree to which an instrument is free from measurement error (Mokkink et al., 2010). By examining reliability, it is possible to assess how well a score represents a true score. Assessing reliability is often measured using intraclass correlation (ICC) for instruments that are scored on a continuous scale. An ICC score represents an estimation of how much the instrument's measurement is free from error. For example, an ICC of 0.7 implies that 70% of the total variation is due to between subject variation rather than measurement error. Although a limitation of assessing reliability is that it does not identify the source of error, different approaches can be used to do this. An inter-rater reliability approach assesses operator error (De Ve et al., 2011), for example where different clinicians measure an individual's blood pressure. This approach also assesses the error from the instrument itself, assuming that all other factors remain constant (operators, an individuals' true score, or other environmental

factors). Intra-rater reliability is normally assessed using a test-retest approach. If the individual remains stable during the test-retest, the reliability coefficient will represent the degree to which the two measurements agree with each other. This can be done by independently repeating the measurement in an individual who has not changed since the last measurement (De Vet et al., 2011). Consideration should be made regarding time intervals between repeated assessments, as an overlapping recall period would introduce bias. Too long a time period may bring about problems with recall, and it may be more difficult to ascertain that the individual remained stable during the test-retest period (Terwee et al., 2010). Reliability represents the variance in a score between individuals due to the true difference in the variable measured (Terwee et al., 2010). Reliability is important when considering comparisons of groups of individuals. The reliability of an instrument reflects how accurately the mean score reflects the 'true' score of the group, without error (Streiner & Norman, 2008). When comparing two groups with different mean scores using instruments with high reliability, the more confidence there is that there is a true difference between the scores, and that this is not only due to error in the instrument's measurement.

(1b) Measurement error

Measurement error is the sum of the systematic and random errors within an individual subject's score in the instrument (De Vet et al., 2011). Reliability is the degree to which an instrument is free from measurement error (Mokkink et al., 2010); measurement error estimates the size of the error within an instrument (Streiner & Norman, 2008).

Measurement error can be estimated by the limits of agreement, as described by a Bland and Altman plot (Bland & Altman, 1986). When assessing the limits of

agreement, systematic error is represented by the mean difference between the first and second scores taken, and random error is 1.96 X SD difference (here SD difference represents the SD of the differences between the first and second measurement). Measurement error is a reflection of an instrument's ability to distinguish an individual's score, and represents the range of scores an individual's 'true' score may be situated in (Streiner & Norman, 2008). For example, in a PA instrument that has small measurement error, only a small change in score is required for the instrument to detect a real change in PA levels. An instrument with large measurement error would require an increased change in score to detect the change as 'real'. This is important when evaluating changes in scores over time in individuals, as it can estimate if changes in scores reflect a true change, or a change that is likely due to measurement error of the instrument used (De Vet et al., 2011).

(1c) Internal consistency

Internal consistency is defined as the extent to which there is inter-relatedness among the items of a uni-dimensional domain of a multi-dimensional instrument. In cases where an instrument needs to be as short as possible, internal consistency can be used (Mokkink et al., 2010). Cronbach's alpha is most commonly used for assessing internal consistency and a Cronbach alpha gives the value of a correlation between items in the domain. Cronbach's alpha is not a measure of uni-dimensionality, as it can show how multiple dimensions are related to each other (de Vet et al., 2011).

(2) Validity

Validity can be defined as an assessment of what degree an instrument measures the variable that it has been constructed to measure (Mokkink et al., 2010). Different forms of validity have been identified with methodologies to analysis those forms within instruments of measurement.

(2a) Face validity

Face validity is a measurement property that gets little attention in research on measurement properties of questionnaires; nevertheless it is a very important assessment of an instrument. Face validity measures the individual items in the instrument and if they adequately reflect the variable they are intended to measure. Important questions to ask about an instrument in face validity are: Is the question asked in a way that would bring about an accurate answer? Does the formula for scoring the questionnaire make sense? Are the questions asked comprehensively? (Terwee et al., 2010).

(2b) Content validity

Content validity is focusing on the items within the instrument and assessing the suitability of those items in reference to what variable the instrument is intended to measure. Content is similar to face validity but is focused on the comprehensiveness of the questionnaire, assessing if all the questions asked are relevant and if all the relevant questions are asked within the questionnaire. Content validity also assesses if there are any questions within the questionnaire

that are irrelevant, or whether questions are scored are appropriately, too coarsely or too finely within the range of scores (De Vet et al., 2011).

(2c) Floor and ceiling effect

Floor and ceiling effect is identified when 15% or greater of responders have scored the lowest possible score or highest possible score. When a questionnaire is used and a floor or ceiling effect occurs, the results affect the reliability coefficient (responders cannot be distinguished). This also affects the responsiveness because responders who score the lowest or highest score cannot change any further in that direction even though this may occur in reality (McHorney & Tarlov, 1995).

(2d) Construct validity

The construct validity comparison of the scores given by a questionnaire compares other instruments that measure the same variable, preferably an already validated instrument. Ideally the comparison instrument would be to a gold standard instrument (criterion validity), in some areas of measurement this is not possible, for example in PA double labelled water may be considered gold standard. There are limitations to double labelled water, as it only gives the total energy expenditure for the day and does not distinguish types, frequency or duration of exercise (Terwee et al., 2010). In construct validity, correlation can be done between the results of the questionnaire and the other identified instruments measuring the same construct (Rennie & Wareham, 1998).

(2e) Criterion validity

When a gold standard measurement is available, criterion validity can be used to assess the reflection the measurement instrument has to that of the gold standard measurement. This can only be assessed if a gold standard measurement exists. Criterion validity is assessed using correlation analysis of the scores from the measurement instrument compared to the gold standard (de Vet et al., 2011).

(3) Responsiveness

The responsiveness of a questionnaire is the ability to detect change over time in the variable being measured; this is an important measurement property for a longitudinal perspective study. Responsiveness of an instrument should be evaluated by comparing a change in score over time compared to changes measured by another validated instrument measuring the same variable (Terwee et al., 2003).

(4) Interpretability

The interpretability of a measurement instrument refers to the ability of a score of an instrument to have a relevant meaning. Measurements in blood pressure have an important clinical meaning (Mokkink et al., 2010), with each unit in the measurement scale indicating a level of risk to hypertension. Interpretability differs from other measurement properties as it does not refer to the quality of the instrument but to the meaning of the score.

4.6 Assessing measurement properties of physical activity questionnaires

Systematic reviews have been published reporting the evidence for appropriate methods and instruments for assessing levels of PA in youth, adult and elderly populations (Chinapaw et al., 2010; Forsen et al., 2010; Helmerhorst et al., 2012; Van Poppel et al., 2010). These systematic reviews found similar conclusions in the measurement properties of self-report PA instruments. The majority of studies included in these systematic reviews showed moderate to strong reliability in test-retest evaluations, particularly in self-report PA instruments for youths aged 18 years (Chinapaw et al., 2010) or below and adult populations aged 18-65 years (Van Poppel et al., 2010). Self-report PA instruments showed lower reliability in adults aged 65 years and over in many of the studies including one of the systematic reviews (Forsen et al., 2010). While youths and adults under 65 years old have shown relatively small measurement error in reporting PA, elderly adults have been reported to have large measurement error in self-reporting PA, leading to a substantial smallest detectable difference and change (Chinapaw et al., 2010; Van Poppel et al., 2010; Forsen et al., 2010). Large measurement error creates issues when measuring PA level for detecting changes over time, as real changes cannot be detected due to large measurement error (De Vet et al., 2011).

All systematic reviews conducted to date have found only moderate correlations between self-report PA instruments and objective measurements (Chinapaw et al., 2010; Forsen et al., 2010; Helmerhorst et al., 2012; Van Poppel et al., 2010). Although there have been studies examining the measurement properties of self-report PA instruments for elderly adults aged 65 years and over, the systematic review by Forsen et al. (2010) concluded that there is a lack of high quality studies

on the measurement properties of PA self-report instruments within this specific population. The findings in these systematic reviews are difficult to interpret for adults with joint pain or OA, as this population is commonly aged 45 years and older. Systematic reviews have generally examined the measurement properties of self-report instruments for adults aged 18-65 and then elderly adults aged 65 and over (Van Poppel et al., 2010; Forsen et al., 2010). It is not clear if the findings in these systematic reviews are applicable to that of adults with joint pain or OA. Overall in systematic reviews in adults and in elderly adults on the measurement properties of self-report PA instruments, reliability is often reported as moderate, although there is high measurement error. Construct validity is also often moderate to other self-report PA instruments but there is poor validity of self-report PA instruments compared to objective measures (Forsen et al., 2010; Helmerhorst et al., 2012; Van Poppel et al., 2010). There would appear to be two promising self-report instruments, with studies reporting high reliability and high correlations to objective measures of PA in adult populations. These are the International Physical Activity Questionnaire (IPAQ) (Craig et al., 2004) and the Physical Activity Scales for the Elderly (PASE) (Washburn et al., 1998).

One systematic review specifically focussed on the measurement of PA in populations with OA has been previously published (Terwee et al., 2011). The review provided an overview of different instruments where there is evidence on measurement properties for assessing PA in populations with OA of the hip and the knee. The review focused on evaluating the evidence of instruments' reliability, responsiveness and validity from previously published literature. The review includes those that are single-item instruments (scales), multi-item instruments (short questionnaires) and objective measures (pedometers) used in a hip and

knee OA population. It was intended that this systematic review would have proposed an instrument for an adult population with knee or hip OA based on the findings. The review concluded that there were not enough high quality studies evaluating measurement properties of PA instruments in OA populations (Terwee et al., 2011).

4.7 Levels of physical activity in osteoarthritis and joint pain

At the time of this thesis, the current level of PA in adult populations with OA or joint pain has not been reported for the UK, or in other countries. Studies have been conducted in other countries examining the level of PA in clinical OA patients including in the United States, Sweden, Germany and Netherlands (de Groot et al., 2008; Rosemann et al., 2008; Dunlop et al., 2011; Dunlop et al., 2011a; Holsgaard-Larsen & Roos, 2012; Felson et al., 2013; White et al., 2013). All of these studies reported differing levels of PA in adults with OA and in each of the studies different measurement tools were used to assess levels of PA making comparisons and estimates of the level of PA in the UK difficult. It would be useful for clinicians and health policy makers to know what the current uptake of PA is at a population level in the UK for those with joint pain. This will allow decision making on how much levels of PA need to increase for adults with OA to be gaining the positive benefits of PA in reduced pain and increased physical function.

In the US using data from large longitudinal cohort studies of populations with OA, a number of researchers have examined the levels of PA in populations with OA (Dunlop et al., 2011; Dunlop et al., 2011a; Felson et al., 2013; White et al., 2013). Dunlop et al. (2011) used data from the Osteoarthritis Initiative (OAI) which is a

multicentre, longitudinal, prospective observational study focussed on knee OA, conducted to observe the development of knee OA and its progression (Dunlop et al., 2011a). This study measured PA using the PA Scales for the Elderly (PASE) (Washburn et al., 1993). The PASE provides a total score ranging from a possible 0-400. A limitation of reporting PA using the PASE is the score as it is not given in units of quantity of PA and it is not possible to compare the level of PA from PASE with recommendations for PA (Washburn et al., 1993). Dunlop et al. (2011) showed that levels of PA can affect physical function. It is not clear from the study if levels of PA in adults also are associated with the overall physical and mental health status of adults with OA or joint pain.

The OAI study also assessed level of PA using the objective PA monitors; ActiGraph GT1M accelerometers (Dunlop et al., 2011a). In 1,111 adults with radiographic knee OA, only 12.9% of males and 7.7% of females achieve recommended weekly levels of PA (Dunlop et al., 2011a) in line with ACSM guideline (Haskell et al., 2007), which match that to PA guidelines in the UK (Bull & Expert Working Groups, 2010). In addition, 40.1% of males and 56.5% of females were found to be classed as inactive, participating in less than one hour of activity each week.

White et al. (2013) reported objective levels of PA in those with radiograph knee OA compared to those without but identified as at high risk of developing knee OA. This study used data from a longitudinal cohort in the United States called the Multicenter Osteoarthritis Study (MOST) (Segal et al., 2013). This was a cross-sectional study that used the StepWatch PA monitor for assessing objective levels of PA in 1,788 study participants with or at risk of having knee OA (White et al.,

2013). This study found that only 6% of males and 5% of females with radiographic knee OA were meeting recommended levels of PA, this was lower than the levels of PA measure in the OAI study, which reported 12.9% of males and 7.7% of females achieving recommended level of PA (Dunlop et al., 2011a). No differences in levels of PA between those with OA compared to adults at risk of OA were found (White et al., 2013). In those with OA, no differences were found in levels of PA in those with mild joint pain compared to those with severe joint pain (White et al., 2013). A limitation for the White et al. (2013) study's conclusion of OA is not associated with levels of PA in that the OA and non-OA comparison group both had very low levels of PA; this caused a floor effect in the measurement tool. A measurement tool with no floor effect could better identify if there is an association between OA and lower levels of PA.

A cross-sectional study of 105 adults in Sweden assessed levels of PA (Holsgaard-Larsen & Roos, 2012), using Sensewear Pro armband activity monitors. Study participants included elderly adults with hip and knee OA, only knee OA and only hip OA and were compared to a non-OA control group. The age range of participants was 65 to 75 years. This study found that those with OA reported lower levels of PA compared to the non-OA control group, 88.7% of the OA population achieved recommended levels of PA and no differences were found between knee and hip OA subgroups (Holsgaard-Larsen & Roos, 2012). This showed a high level of PA in comparison to the other studies in adults with OA (Dunlop et al., 2011a; White et al., 2013). This suggests that levels of PA may vary in different countries, as this study found a much higher level of PA measured objectively to that of the studies in the United States, although different types of activity monitors were used in the studies. The limitation of this study was that the

Sensewear Pro armband activity monitors have since been shown to overestimate levels of PA in adults with hip OA (Hernmann et al., 2014).

A study in the Netherlands compared levels of PA using AM activity monitors in those with severe hip or knee OA awaiting total joint replacement with healthy controls (de Groot et al., 2008). They found that OA has a clinically important impact on lowering levels of PA. In contrast, an American study found that low levels of PA were in both OA and non-OA populations with no statistical association between lower levels of PA and joint pain (White et al., 2013). It is unknown if the levels of PA in UK adults with joint pain are similar to the general population without joint pain or whether they are at risk of lower levels of PA.

In adults with OA, the International PA Questionnaire (IPAQ) has also been used to assess levels of PA (Rosemann et al., 2008). A cross sectional study using data from a large scale general practice survey on OA patient outcomes was conducted in Germany (Rosemann et al., 2008). The study found that 38% of knee OA or hip OA populations were sufficiently active to ACSM guidelines, in addition, 7-10% of that population were highly active. The levels of PA reported in this study were much higher compared to the objective measures of PA in studies from the United States (Rosemann et al., 2008). Presence of knee or hip OA were associated with lower levels of PA compared to age-matched controls with no knee or hip OA. Those with lower levels of PA showed decreased levels in physical health status, mental health status and greater in OA symptoms, measured by the AIMS-2 short form questionnaire on health status (Rosemann et al., 2008). This suggests that more severe OA symptoms could be contributing to

lowering PA levels or those with sedentary lifestyles are associated with more severe OA.

Studies of PA in populations with OA would suggest that overall levels are lower compared to healthy, non-OA comparison groups (de Groot et al., 2008; Holsgaard-Larsen & Roos, 2012). This has not been investigated in the UK, in fact one study contradicts this hypothesis (White et al., 2013). The level of PA would seem to vary with percentages of the population with OA being physically active to guidelines ranging from 5% (White et al., 2013) to 88.7% (Holsgaard-Larsen & Roos, 2012). The level of PA reported would seem to depend on the population and the measurement instrument. There is currently no evidence of the validity of the PA monitors used in any of the studies investigating OA populations (Dunlop et al., 2011a; Holsgaard-Larsen & Roos, 2012; White et al., 2013). The variance across different countries in levels of PA for adults with OA suggests that the levels of PA in adults in the UK with joint pain may differ from other previous studies. It would be preferable if a uniformed method for measuring PA in adults with OA or joint pain was established (De Vet et al., 2011). Within the UK, the levels of PA in adults aged 45 years and over with joint pain or OA is still unknown and the best method for measuring PA in this population is not clear (Terwee et al., 2011). Measuring levels of PA in adults with and without joint pain in the UK would allow for understanding of how much focus on increasing levels of PA in this population is required and establish if interventions for raising PA in UK adults aged 45 years and over with OA is required. Measuring levels of PA can also be used to assess the association of PA and presence of joint pain, comparing if adults with different sites of joint pain are at risk of lower levels of PA. Measuring PA in adults aged 45 years and over with OA allows for investigating association

between levels of PA and other outcome measures to gain better knowledge of the interactions between PA levels, OA and health status. Selecting the best measurement instrument of PA will depend upon the measurement properties and the evidence for how appropriate the instruments are in adult populations with OA.

4.8 Summary

This chapter evaluated the possible approaches for measuring PA levels in adults aged 45 years and over and determined that self-report instruments would seem the most appropriate for use in the target population. It would not be practical to use objective measures of PA in all joint pain research due to the practical implications and high costs with large sample sizes. There are large numbers of self-report PA instruments available; selecting the most appropriate depends on the measurement properties of the instruments in adults aged 45 years and over with joint pain. The limitations of self-report PA instruments are the precision of measurement and responsiveness over time, assessing the instrument's measurement properties would allow for selecting the one with the highest precision of measurement.

This chapter has explored a rationale for selecting self-report instruments as an appropriate methodology to measure levels of PA in adults with OA or joint pain. The main findings of this chapter were:

- The most appropriate approach for measuring PA in population level research is by using self-report instruments

- A theoretical framework assessing measurement will allow for identifying the most appropriate self-report instrument using classical test theory

- Despite a number of systematic reviews on the measurement properties of self-report PA instruments it is not yet clear which is the most appropriate for adults with joint pain or OA

- Levels of PA in OA populations are heterogeneous depending where the study was conducted and the measurement instrument used.

Chapter five: Methods

5.1 Introduction

This chapter presents the rationale for the thesis and outlines the methods that were used to achieve the three aims. The aims of this thesis are:

1. To describe the self-reported levels of PA of UK adults aged 45 years and over

2. To evaluate the measurement properties of reproducible self-report PA instruments in adults aged 45 years and over with and without joint pain or OA

3. To evaluate the measurement properties of the International PA Questionnaire Short Form (IPAQ-SF) and the PA Scale for the Elderly (PASE) in adults aged 45 years and over with joint pain

This chapter provides a description of:

- Cross-sectional observational and population survey methodology
- Systematic review methods
- Evaluation in measurement properties of self-report measures
- The generic statistical methods used in this thesis

5.2 Cross-sectional and population survey methods

5.2.1 Cross-sectional studies

To achieve aim one of this thesis a cross-sectional self-report population survey was used. Cross-sectional studies can measure a sample at a single time-point (Silman & Macfarlane, 2002). Cross-sectional studies are used to describe the characteristics of the sample and to estimate parameters in this population by taking a sub-sample (Fletcher et al., 2013). Cross-sectional survey methods are typically descriptive rather than analytical or experimental. They are commonly used in health research to describe the current health status of populations. This can be conducted by makeing comparisons of health status in two or more different populations or to show associations in parameters (Bland, 2000; Bowling, 2009; Bryman, 2012). This method cannot be used to draw firm conclusions as there is not sufficient theoretical rationale for a cause and effect relationship (Rothman et al., 2008). Instead observations on occurrences in phenomenon within a population, such as prevalence and incidence can be described (Silman & Macfarlane, 2002).

Investigations of associations between subgroups for a particular phenomenon are also possible in cross-sectional studies (Fletcher et al., 2013). When using survey methods, the relatively short time taken for data collection has the advantage over experimental and longitudinal studies that require data collection at multiple time-points and follow up, which can usually run from 12 months to any number of years (Hennekens et al., 1987; Bowling, 2009). Compared to these studies, cross-sectional studies take a relatively short time to report findings (Bryman, 2012).

Cross-sectional studies use smaller samples of populations of interest to estimate relationships between variables. They can be used to determine the prevalence of disease and if there is an association between an exposure and disease state (Silman & Macfarlane, 2002). Consideration needs to be given to the existence of confounding factors when describing relationships (Groenwold et al., 2008). One strength of cross-sectional studies is the smaller cost. The weakness of cross-sectional studies is that the findings can be prone to confounding of observed relationships (Silman & Macfarlane, 2002).

5.2.2 Population surveys

To address the first aim of this thesis, cross-sectional research methods are the most appropriate approach. The advantages of this approach for collecting data were:

1) It was possible to distribute the survey by post to a large population.

2) A number of different outcome measures were included in one survey (Silman & Macfarlane, 2002).

3) Limitations of the risk of bias can also be minimised in the cross-sectional study.

The limitations of a self-report population survey are: risk of low response rate, incomplete questionnaires or miss-interpretation of questionnaires and recall bias (Fowler, 2014). Limitations can be addressed by identifying the measurement properties of instruments included in the study; using potential confounders during

statistical analysis and evaluating non-responders and responders in differences in characteristics.

Population surveys are vulnerable to bias from non-responders when response rates are low (Fowler, 2014). This bias is caused when there are systematic differences in the characteristics of those who responded to the survey and those who did not (Bowling, 2008). Non-responders also reduce the sample size of the survey, therefore losing precision. Differences in responders and non-responders can lead to error in study conclusions (Silman & Macfarlane, 2002). It is difficult to get information on non-responders due to the ethical considerations, unless consent to demographic information was given (Bowling, 2008).

A limitation of retrospective self-report surveys is that they can include recall bias. Recall bias in survey research arises when a misclassification of a retrospective question is made due the respondent recalling events/symptoms or behaviour differently from the 'true' situation (Coughlin, 1990). Recall bias can cause incorrect shifts away from and towards the null hypothesis. Well-designed surveys can minimise recall bias by motivating responders to correctly recall responses to questions (Coughlin, 1990), this will depend on the responders and the measurement properties of the instruments used in the questionnaire.

A causal relationship between two variables can be explained by a third variable, which is independent but associated with the other two, this is known as a confounding variable (Grimes & Schulz, 2002). If the confounding variable is not observed there is risk of concluding incorrect relationships between variables. Potential confounding variables can be accounted for and controlled to estimate the 'true' relationship between two variables when excluding the effect of the

confounding variables (Silman & Macfarlane, 2002). An example of a confounding variable is age in the relationship between income and risk of cancer in US adults (MacKinnon et al., 2000). Adults aged 45 years and over are likely to have a high income as they have been in the workforce longer compared to younger adults, and adults aged 45 years and over are at higher risk of cancer compared to younger adults. The relationship between income and cancer is related through age (MacKinnon et al., 2000).

Despite possible limitations, the most appropriate research method design for aim one is the use of a self-report cross-sectional survey, as it will allow for estimations generated regarding the level of PA in a UK adult population with and without joint pain.

5.3 Systematic review methods

To address aim two, a systematic review was selected to evaluate measurement properties of self-report PA instruments. Systematic reviews aim to identify, evaluate and summarise the findings of all relevant studies within an area of interest (Higgins & Green, 2008). A systematic review allows for an overview of all self-report PA instruments currently available and the evaluation of their measurement properties. Individual studies can contain forms of bias, flawed methodologically or be limited in other ways leading to its findings being misinterpreted. Systematic reviews can, where appropriate, pool these individual studies together and present findings in a more reliable and precise way. The methodology of a systematic review aims to reduce bias or methodological flaws by being designed in a scientific and reproducible way (University of York Centre for Reviews, and Dissemination, 2009).

A systematic review allows for a more objective evaluation of the evidence compared to non-systematic and can contribute to summarise findings of all related original research (Higgins & Green, 2008). If conducted with adequate methodological rigour, a systematic review can identify risk of bias. The review process of systematic reviews can be fully reported, making the protocol of searching and synthesising of evidence, fully transparent and easy to directly replicate.

Development of systematic review methods in health have largely been concerned with searching literature and synthesising findings from RCTS which have tested new interventions for the benefits of particular health outcomes (University of York Centre for Reviews, and Dissemination, 2009). The systematic review conducted for aim two of this thesis was undertaken to examine the measurement properties of self-report instruments that assess levels of PA. Therefore, this the systematic review will differ in its methodological approach compared to systematic reviews on RCTs or epidemiological studies (De Vet et al., 2011).

5.3.1 Search strategy

In recent years there has been research and development into the assessment of measurement tools using a systematic review. Research has examined the methods for valid search strategies of literature in electronic databases (Terwee et al., 2009), data extraction (Terwee et al., 2010) and quality assessment of articles (Mokkink et al., 2010) that assess measurement properties of instruments. The advantage of a standardised search strategy, data extraction tools and quality assessments forms is that they minimise the risk of reviewer bias and therefore reduce reporting bias.

5.3.2 Quality assessment tools

A systematic review will still have limitations in the original studies that may create bias in the results. It is therefore necessary to assess articles selected within the review by conducting a quality assessment (University of York Centre for Reviews, and Dissemination, 2009).

A quality assessment of papers aims to establish if the findings of a study can be identified as a ‘true’ representation of a target population of interest. This is achieved by considering factors of a study such as: appropriateness of the study design, risk of different forms of bias, quality of reporting findings and use of statistical analysis (University of York Centre for Reviews, and Dissemination, 2009).

Systematic reviews should be explicit in the methods used to quality assess the included studies (University of York Centre for Reviews, and Dissemination, 2009). Various quality assessment tools for systematic reviews have been developed to standardise the process (University of York Centre for Reviews, and Dissemination, 2009). The different tools for quality assessments vary depending on what kind of systematic review they are assessing (University of York Centre for Reviews, and Dissemination, 2009). An example of quality assessment tools are the PRIMSA criteria (Preferred Reporting Items for Systematic reviews and Meta-Analyses) for RCTS in health research (Moher et al., 2009) or the COnsensus-based Standards for the Selection of health Measurement INstruments also known as the COSMIN checklist (Mokkink et al., 2010). As the systematic review for aim two is focussed on evaluating instrument measurement properties, the COSMIN checklist was selected for quality assessment.

The COSMIN checklist was constructed for the purpose of helping researchers and healthcare professionals improve the selection of health assessment instruments. The checklist aims to give an assessment of not only the instrument but also the studies that evaluate the instrument (Mokkink et al., 2010).

The construction of the COSMIN checklist came from a Delphi study to find a consensus on different measurement properties across researchers and healthcare professionals (Mokkink et al., 2010). An international group of researchers and experts were involved in the final agreement in definitions and method of assessing measurement properties (Mokkink et al., 2010). The construction and testing of its reliability has been published in a number of articles (Mokkink et al., 2010).

The COSMIN checklist provides a comprehensive evaluation of studies assessing measurement properties of instruments. A quantitative version of the checklist is available allowing for scores of studies to be compared. Problems can occur with the COSMIN as the checklist asks for specific detail regarding statistical analysis and methods that few studies often conduct, leading to many questions not being applicable to manuscripts (Mokkink et al., 2010). Reliability and levels of agreement of the inter-rater use of the COSMIN checklist are high with 68% of items in the list with at least 80% agreement (Mokkink et al., 2010). Reliability was low, making higher quality studies difficult to differentiate from lower quality studies (Mokkink et al., 2010). These Kappa coefficients represented acceptable levels of agreement but reliability was low between different raters using COSMIN possibly because of the difference in experience of raters in conducting systematic reviews which caused differences in the scoring (Mokkink et al., 2010).

5.3.3 Data extraction

Much like the quality assessment of articles in systematic review data extraction, a systematic process can limit the chance of reviewer bias. Prior to conducting a systematic review, consideration should be given to the data that will be collected from the included studies; data collection should reflect the research question and aim of the systematic review. Forms and tables can then be constructed so that each article has information collected in the same way, which will reduce risk of reviewer bias (University of York Centre for Reviews, and Dissemination, 2009).

Standardised data extraction forms and protocols have been developed further aiding systematic reviewers in data extraction within specific systematic reviewing areas (University of York Centre for Reviews, and Dissemination, 2009). An example of a data extraction checklist highly relevant to this thesis is the Quality Assessment of PA Questionnaire (QAPAQ) (Terwee et al., 2010).

The QAPAQ checklist was developed for the assessment of measurement properties in self-reported PA questionnaires (Terwee et al., 2010). The aim of the QAPAQ is to allow for the most appropriate PA questionnaire to be selected for a certain purpose or setting. The checklist was developed using previous literature on qualitative attributes of questionnaires (Terwee et al., 2010). Data from the COSMIN Delphi study were also used in the QAPAQ construction (Terwee et al., 2010). The QAPAQ is split into two sections: section one, appraising the qualitative attributes of the PA questionnaire, such as setting of PA and the target population, and section two, appraising the measurement properties of the PA questionnaire.

In terms of the qualitative attributes of the questionnaire, the QAPAQ criteria suggest that a questionnaire should contain a clear description of what the questionnaire is intended to measure, who the questionnaire is meant for and its purpose (Terwee et al., 2010).

The QAPAQ also identifies general study properties such as the study population, methods used and statistical analysis, such as the statistical details of the measurement properties (Terwee et al., 2010).

The COSMIN checklist and the QAPAQ checklist are not the only available checklists for systematic reviews examining the assessment of measurement properties of outcome measures. An Oxford University based research group has, for many years, conducted research in the area of health-related self-reported outcome measures for patients with various different health conditions (Smith et al., 2005). Much of the published work from this research group includes reports for the Department of Health on the use of health-related self-reported outcome measures in patients with specific health conditions (Smith et al., 2005).

The Oxford based research group developed a standardised form for use with studies that examine the measurement properties of instruments, to aid the quality assessment and data extraction process for a systematic review (Smith et al., 2005). Unlike the COSMIN checklist, there is little published work with reference to this form, it was developed by a group of experts within the field of measurement and is an appropriate method particularly in the area of patient self-reported outcomes (Smith et al., 2005).

In this thesis various approaches for a search strategy, quality assessment and data extraction were considered and are described. In brief, the search strategy outlined by Terwee et al. (2009), the COSMIN checklist (Mokkink et al., 2010) and the QAPAQ checklist (Terwee et al., 2010), were chosen for this thesis as both are standardised approaches using definitions that have been established by Delphi studies and agreed by experts in measurement property evaluations (Terwee et al., 2009; Mokkink et al., 2010; Terwee et al., 2010)

5.4 Evaluation in measurement properties of self-report measures

De Vet et al. (2011) identified four key elements when assessing measurement properties of self-report instruments:

1. Identifying the construct of interest, or the name(s) measurement instruments for evaluation
2. Identifying the population of interest
3. Identifying the type of instrument of interest (performance based, self-report, objective, imaging or interviewing)
4. Evaluating measurement properties in the instruments of interest

The construct of interest in this thesis is measuring the quantity of PA that individuals will participate. The frequency, duration and intensity of PA are important and useful for interpreting the clinical benefit and comparing to PA guidelines (Bull & Expert Working Groups, 2010).

Considering which measurement properties to evaluate depends upon the construct that is measured and the purpose of the measurement. In this thesis the measurement properties that were chosen for evaluation were: reliability, standard error of measurement, constructs validity and responsiveness. Chapter four (page 54) includes a definition and approaches to evaluate these measurement properties. All of the measurement properties outlined by the COSMIN were used to evaluate the precision of the instruments' and responsiveness was selected to evaluate instruments' ability to measure change in PA levels over time.

5.5 Overview of the generic statistical methods used in this thesis

In this thesis statistical analysis was conducted using the software package Statistical Package for the Social Sciences (SPSS) (IBM Corp. Released 2013. IBM SPSS Statistics for Windows, Version 22.0), unless otherwise stated in the analysis chapters. Specific statistical approaches used are described within the relevant chapters and sections of this thesis, this section will describe some statistical approaches used in the description of study participants. Averages for continuous variables are presented in mean with variance from the means presented in standard deviations. For data that were not normally distributed, median and interquartile ranges were used. For categorical data, frequency and percentages of proportions were displayed. For comparisons of independent groups; chi-squared, independent t-test and one-way analysis of variance were conducted when appropriate. Where data was not normally distributed, Wilcoxon rank sum tests and Kruskal Wallis tests were conducted. When analysing paired data of single groups in repeated measurements, NcNemar test or paired t-tests were suitably conducted. Associations between variables were assessed using

correlations and Spearman’s rank coefficient. In this thesis where all statistical models were conducted, data was first assessed for in the statistical model’s data assumptions, this was conducted by following the guidance of statistical textbooks (Pallant, 2011; Field, 2012).

5.6 Summary

This chapter has described and explained the selection of methodological approaches used to achieve the three aims of the thesis. The next chapter will build on the knowledge in this chapter and shall provide:

- The protocol of the Management of Osteoarthritis in Consultations Study (MOSAICS) population survey and consultation questionnaire

- Describe the participants inclusion and exclusion crtieria for the MOSAICS population survey and consultation questionnaire and the outcome measures selected from MOSAICS for this thesis

Chapter six: The Management of Osteoarthritis in Consultations Study (MOSAICS) – Secondary data analysis

6.1 Introduction

Chapter five provided a discussion on the strengths and limitations of the methods used in this thesis. This chapter describes an overview of the Management of Osteoarthritis in Consultations Study (MOSAICS), which this thesis used data from to achieve the thesis aims. The overview of the MOSAICS in the chapter includes:

- A description of the structure of MOSAICS
- Eligibility criteria of participants of MOSAICS
- A description of the outcome measures used in the MOSAICS population survey and the consultation questionnaires.

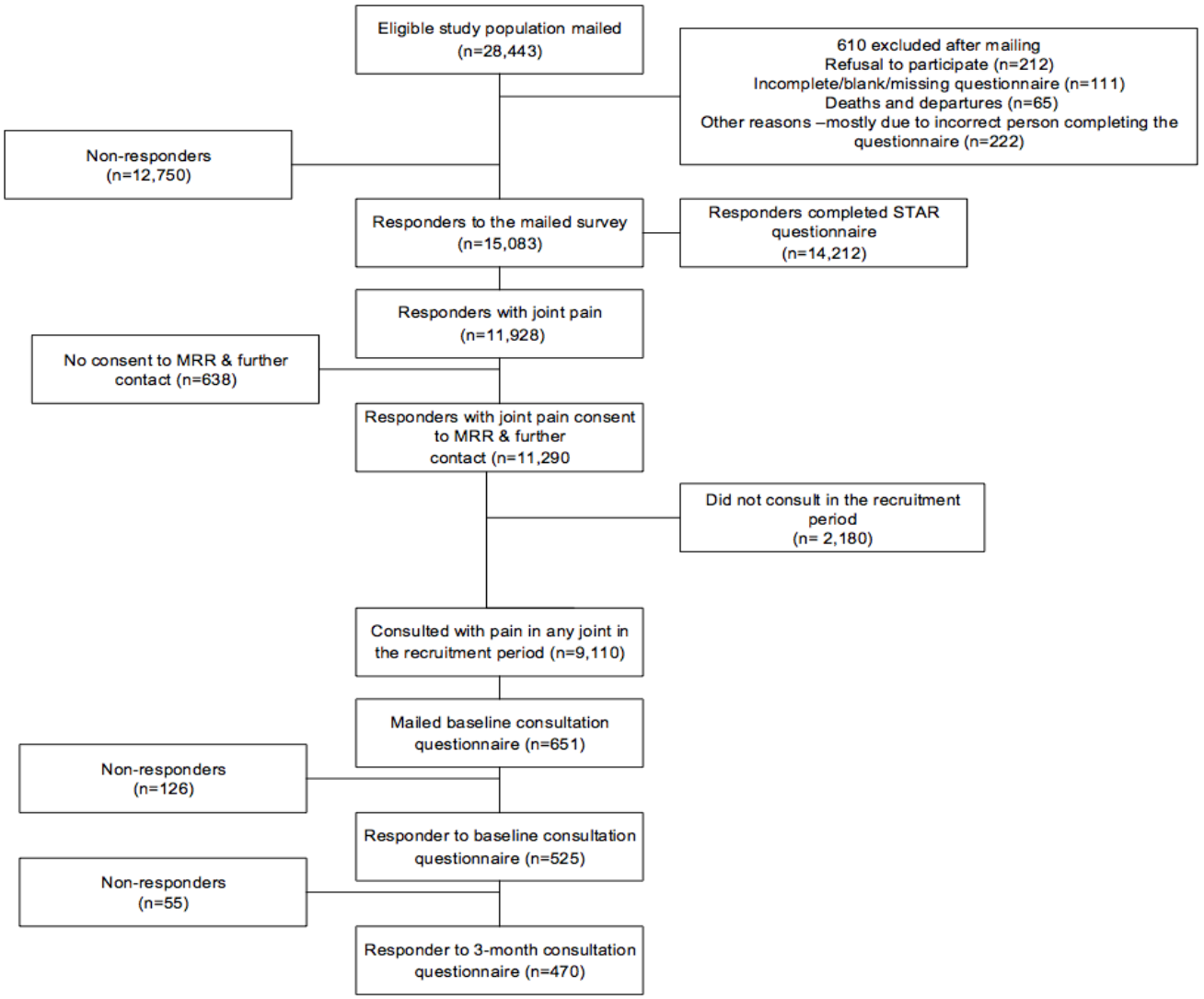
6.2 Overview of the MOSAICS study

The protocol paper of the MOSAICS study is available in the thesis appendix (Appendix 1.1). The MOSAICS study was designed to investigate the clinical and cost effectiveness together with the feasibility for implementing the NICE core treatments for OA in a UK primary care setting and the impact on self-management in adults aged 45 years and over (Dziedzic et al., 2014). This was conducted based on evidence for a model primary care consultation which was drawn from the findings of two Dephi consensus studies (Finney et al., 2013; Porcheret et al., 2013). The primary aim of the MOSAICS project was to: determine the clinical and cost effectiveness of using a model primary care

intervention in patients with OA. Further details on MOSAICS can be found in the study protocol (Dziedzic et al., 2014).

The MOSAICS study used mixed methods, within a cluster randomised controlled trial (RCT). Two components of MOSAICS were used in this thesis: the population survey and the consultation questionnaires. Figure 6.1 displays the flowchart of the MOSAICS population survey and the consultation questionnaire. The design of the MOSAICS study was agreed upon, prior to thesis work commencing. There was no input of the findings of this thesis into the design of the MOSAICS study. The systematic review in Chapter Eight of this thesis was conducted after the outcome measures of PA were selected for MOSAICS, and did not inform MOSAICS.

Figure 6.1 Flowchart of MOSAICS population survey and consultation questionnaire



The study setting for MOSAICS was eight general practices in Central England; eligibility criteria for practices in the MOSAICS study are displayed in table 6.1 and for patients in table 6.2. This eligibility criteria was for the RCT component of MOSAICS, this was also the setting and the participants in which the MOSAICS population survey was conducted. The study was approved by the North West 1 Research Ethics Committee, Cheshire with the research committee reference number 10/H1017/76 (Appendix 1.2), and the International Standard Serial Number for the MOSAICS study is ISRCTN06984617. Health care professionals and participants gave informed consent to participate in the study (Dziedzic et al., 2014).

Table 6.1 Eligibility criteria of general practices in MOSAICS study

	Eligibility criteria
Inclusion criteria	Member of the Central England Primary Care Research Network or a Keele Research Network Practice
	At least two GPs willing to undertake the study as per protocol, for example act as a control or intervention practice
	Willing, and able, to allow one or two of their practice nurses to be trained
	Able to physically accommodate the nurse clinics in the practice
	Uses the EMIS computerised consultation system
	Nurses and GPs consenting to follow-up by the MOSAICS study team
	GPs willing to be trained
	Nurses willing to be trained
	Nurses who consent to being observed and audio recorded in clinics

Table 6.2 Eligibility criteria of patients in MOSAICS study

	Eligibility criteria
Inclusion criteria	Males and Females
	45 years and over
	Registered with a MOSAICS study practice
	Reported joint pain in the population survey and consented to further contact from the study team and medical record review
Exclusion criteria	Excluded via GP screen of practice list
	Unable to give fully informed consent, for example, learning difficulties or dementia
	Resident in a care or nursing home
	History of serious disease, for example, malignancy, terminal illness
	Unable to consult in the general practice
	Flagged as excluded from research in that practice

6.3 MOSAICS population survey

For the first aim of this thesis, data analysis of the MOSAICS population survey was undertaken. The population survey was a cross-sectional survey mailed out to a sample of 28,443 adults aged 45 years and over registered in the eight participating primary care practices. Prior to mailing, the practice was given the opportunity to screen and exclude participants, for example, having psychiatric illness, or having experienced a recent family bereavement.

The population survey was distributed in a two stage mailing out system. Individuals were excluded if they were considered ineligible by their general practice or contacted the research team themselves wishing to take no further part in the study. In the first stage, eligible participants were sent a survey along with an invitation to take part in the study and information about the MOAICS study. In stage two after two weeks, non-responders were sent a reminder survey letter.

Data management of the MOSAICS population survey was conducted by the primary care research network. This included checking study ID numbers to GP practice database to ensure the correct individual was sent a survey. Consent, opting out and potential participants that were ineligible were manually managed by the clinical trials unit administration staff.

Returned surveys were then inputted using a teleform machine method using a teleform reader and verifier (Cardiff Software Inc). This method automatically detects responses to items and records them in an electronic database and is a valid method for data entry (Jinks et al., 2003). A member of the MOSAICS research administration team then performed a 1 in 10 check of the electronic database to assess data quality. Data cleaning was also conducted by the MOSAICS lead statistician (Dr. Martyn Lewis) to ensure data quality and to locate if there were any instances of outliers within the data that may have been incorrectly inputted. The statistician also carried out data cleaning for missing data or anomalies.

Data collected in the MOSAICS population survey included respondents' demographics and participants' reported outcome measures. This thesis did not use all of the outcome measure in the MOSAICS population survey as they were not applicable to 'aim one'. The protocol article includes a description of all outcome measures used in the MOSAICS population survey (Dziedzic et al., 2014). The data collected to achieve aim one of this thesis are summarised in table 6.3 and described in sections 6.3.1 and 6.3.2. The method used for scoring the measures is also reported.

Table 6.3 Data collected in MOSAICS population survey used in this thesis

Demographic measures	Age (years)
	Gender (male or female)
	Body mass index (BMI, kg/m^2)
	Postcode (English indices of deprivation score (IMD)) (Department of British Communities and Local Government, 2012)
Participant reported outcomes	General Health Status (SF-12) (Ware et al., 1996)
	Joint pain location over the last 12 months (hip, knee, hand and/or foot)
	Numerical rating scale (NRS) for joint pain intensity during the past 12 months (Keller et al., 2004)
	Level of PA (short telephone activity rating (STAR) questionnaire (Matthews et al., 2005)

6.3.1 Demographic measures

(1) Body Mass Index (BMI)

BMI was calculated using the standardised calculation kg/m^2 using self-report height and weight. This thesis also categorised responders into BMI categories using the NICE guidelines for BMI health status: under-weight BMI (<18.5kg/m^2), normal range BMI (18.6-24.99kg/m^2), overweight BMI (25-29.9kg/m^2), obese BMI class I (≥30kg/m^2) (NICE, 2006). BMI categories can further indicate the health status of the population in this study; BMI is not only an important factor in an individual's health but is also associated with development and progression of knee OA (Reijman et al., 2007) and joint pain in the lower limbs (Adamson et al., 2006).

(2) English Indices of Deprivation Score (IMD)

Socio-economic status of the study population was determined based on the participants' postcode and the Index of Multiple Deprivation (IMD) (Department of British Communities and Local Government, 2012). The IMD is calculated by a UK government department based on the areas average income, employment, health markers, education, living environment and crime for different regions of the UK (Department of British Communities and Local Government, 2012). Responders were given an IMD score based on their postcode and that corresponding postcode's IMD value. A lower scoring IMD value indicates a higher level of socio-economic deprivation, although it should be noted this is not a socio-economic score of an individual but based on the area that individual is living in.

6.3.2 Participant reported outcomes

(1) General Health Status, The 12-item short form health survey (SF-12)

The SF-12 was constructed using statistical data reduction of the 36-item health survey while still reproducing the same physical component scores (PCS) and mental component scores (MCS) as an indicator for an individual's health status (Ware et al., 1996). The SF-12 was developed from a longer instrument of general health the SF-36 in a general adult population in the United States (Ware et al., 1996). It has since translated into different languages and has been used in different populations in a wide variety of populations, including the UK (Gandek & Ware, 1998).

The SF-12 has been previously shown to be valid when compared with other health status instruments and has shown to be reliable with small measurement

error in the UK general population (Gandek et al., 1998). In an OA sample of n=651 participants, the SF-12 was shown to have good structural validity for both the PCS and MCS (Gandi et al., 2001). It also demonstrated no floor and ceiling effect, low levels of missing data and significant correlations with relevant clinical outcomes as a measure of construct validity (Gandi et al., 2001)

For the MOSAICS study, version 2 of the SF-12 was used. Version 2 of the SF-12 consists of 12 items, with 10 items having a 1-5 scale and 2 items of activities having a 1-3 scale. This scoring method allows for computation of missing data and has shown to be valid without losing data (Ware & Gandek, 1998). Scoring of the SF-12 PCS and MCS was carried out using the procedure outlined in the version two SF-12 scoring document (Ware et al., 1996). Scores were standardised to an adult population from the United States with a mean of 50 and standard deviation of 10 in the PCS and MCS. Eight domains of the SF-12 were scored, these domains consist of: physical functioning, role-physical, bodily pain, general health, vitality, social functioning, role-emotional and mental health through which PCS and MCS were calculated. In a survey of 1751 adults in the UK aged 18 year and over a mean PCS of 50.8±9.4 and a mean MCS of 52.1±8.7 (Gandek et al., 1998).

(2) Pain location and pain intensity

Responders were asked if they had experienced any pain in or around the hip, knee, hand and foot in the last 12 months (see appendix 1.3). The hip, knee, hand and foot were selected to correspond with the NICE guidelines on management of OA (NICE, 2014). These pain questions were taken from a previous epidemiological study on the prevalence of pain, The North Staffordshire

Osteoarthritis Project (NorStOP) and examined pain in the last 4 weeks (Thomas et al., 2004). Responders were asked to report on average how intense their pain was in each site and also during the past twelve months using a 0-10 NRS for pain (see appendix 1.4). This 0-10 rating scale for pain has previously been validated in an adult population with arthritis (Keller et al., 2004). The NRS has demonstrated high reliability and construct validity in comparison to other measures of pain, including the WOMAC pain intensity score and visual analogue scale of pain (Keller et al., 2004). In responders with more than one painful joint, the highest scoring joint was taken (Croft et al., 2010).

Pain intensity categories were also created using a method previously developed and validity tested by Zelman et al. (2005). Three pain categories were derived: 0 - no pain, 1-4 - mild pain and 5-10 - moderate/severe pain intensity (Zelman et al., 2005). These categories have been previously used in OA as pain intensity categories (Kapstad et al., 2008).

(3) Short Telephone Activity Rating (STAR) questionnaire

The STAR questionnaire is a short, three-item self-report measure of levels of PA. It was developed originally to be used as a short telephone interview instrument to assess the level of PA in an adult population for research purposes (Matthews et al., 2005). It has since been used in postal surveys for responders to self-report their level of PA (Holden et al., 2012). An expert group of PA and sports medicine researchers designed the STAR based on previously designed self-report instruments of PA: the IPAQ and PASE (Washburn et al., 1993; Craig et al., 2003). Two versions of the STAR questionnaire were developed: an open response version and closed response version. The STAR questionnaire showed to be

reliable, in a test-retest study using the closed STAR in a general adult population consistency of classification was a kappa of 0.76 (Matthews et al., 2005). Construct validity of the STAR had demonstrated as moderate compared to a 24 item PA questionnaire (24PAR) and low compared to Actigraph PA monitor (Matthews et al., 2005). Further evaluation of the measurement properties of the STAR questionnaire can be found in the systematic review of this thesis in chapter nine. The MOSAICS population survey used the closed version to standardise responses and for data extraction using a teleform machine. Items one and two of the STAR ask responders to indicate the frequency and duration of participating in moderate PA over a usual week respectively. The third item asks responders to indicate the frequency in which they participate in vigorous activities for at least 20 minutes in a usual week (Matthews et al., 2005).

The STAR questionnaire was scored corresponding to recommended guidelines in levels of weekly PA (Bull & Expert working panel, 2010) and using guidelines for scoring the STAR questionnaire (Matthews et al., 2005). The STAR can be scored to categorise responders into three levels of PA: inactive, somewhat active and active to DH recommendations (Bull & Expert working panel, 2010). Table 6.4 describes the STAR questionnaire categories.

Table 6.4 Description of categories in the STAR questionnaire

Category	Description
Inactive	where responders report participating in moderate levels PA (3-6 metabolic equivalents, METS) less than once a week and report participating in vigorous levels of PA (>6 METS) less than once a week.
Somewhat active	where responders report participating in moderate PA five or more times a week lasting at least 30 minutes each day or participating in vigorous levels of PA at least three times a week.
Active	where responders report participating in moderate PA five or more times a week lasting at least 30 minutes each day or participating in vigorous levels of PA at least three times a week.

6.4 MOSAICS consultation questionnaire

Responders to the MOSAICS population survey were asked for consent to their medical records being reviewed and for further contact for the MOSAICS consultation questionnaires. If a responder consulted their general practitioner for joint pain they were identified via fortnightly electronic searches of consultation data, those identified were then sent a baseline consultation questionnaire.

For the MOSAICS baseline and three month consultation questionnaire, a three stage mailing process was conducted in line with the established procedures of the Arthritis Research UK Primary Care Centre. Stage one, the participants were sent information about the study along with a baseline consultation questionnaire (Dziedzic et al., 2014). Stage two, included sending a reminder post card to those who had not returned the consultation questionnaire within two weeks. Finally, in stage three after a further two weeks, non-responders were sent an additional questionnaire and a reminder letter. For three month follow up consultation, a

questionnaire was then mailed out to participants who had replied to the baseline consultation questionnaire (Dziedzic et al., 2014) using the same procedure.

The MOSAICS baseline and three month consultation questionnaires instruments and demographic data that were used in this thesis are described in table 6.5. The IPAQ-SF and PASE were included in the baseline and three month consultation questionnaire specifically for the purpose of meeting aim three of this thesis.

Table 6.5 Data collected in MOSAICS baseline and three month consultation questionnaire used in this thesis

Demographic measures	Age (years)
	Gender (male or female)
	Body mass index (BMI, kg/m^2)
	Postcode (English indices of deprivation score (IMD)) (Department of British Communities and Local Government, 2012)
Participant reported outcomes	General Health Status (SF-12) (Ware et al., 1996)
	The EuroQoL five dimensions, 3 levels (EQ-5D 3L) (Rabin & Charro, 2001)
	Joint pain location over the last 3 months (hip, knee, hand and/or foot)
	Numerical rating scale (NRS) for joint pain intensity during the past 3 months (Keller et al., 2004)
	International Physical Activity Questionnaire Short Form (IPAQ-SF) (Craig et al., 2003)
	Physical Activity Scales for the Elderly (PASE) (Washburn et al., 1993)
	Uptake of exercise and PA over the last 3 months (NICE, 2014)
	Global perceived change in joint problems

The measures of BMI, IMD score, SF-12, joint pain, joint pain intensity and uptake of exercise and PA are described above in the description of the population survey, only the time frames of the joint pain, joint pain intensity, uptake of PA and exercise were changed at three months. The sections below describe the EQ-5D, the global perceived change in joint problems, the IPAQ-SF and the PASE.

(1) EQ-5D

The EQ-5D 3L was developed by an international research group (EuroQOL) (Rabin & Charro, 2001). It is a simple and generic patient reported outcome measure of health status. The rationale for the EQ-5D and SF-12 is that the EQ-5D includes a single index score which can be used for clinical and health-economic evaluations (Rabin & Charro, 2001). The EQ-5D has five items on five domains of general health status: mobility, self-care, usual activities, pain/discomfort and anxiety/ depression. Each item has a 1-3 rating scale, depending on responder's health state in each domain (Rabin & Charro, 2001). The EQ-5D has been previously validated against other measures of health status, including the SF-36, showing high agreement between both instruments in a study sample from the 1980 UK general adult population (Brazier et al., 1993). In a study with 208 Dutch community dwelling adults, the EQ-5D had shown to be reliable in a test–retest study (Van Agt et al., 1994). The EQ-5D has also been shown to have good scores in terms of reliability and validity for an adult population with knee OA (Fransen & Edmonds, 1999). In the evaluation of the EQ-5D in adults with knee OA, acceptable reliability was found in a non-parametric analysis and moderate correlations to the SF-36 were also found, although there were wide 95% CI in these correlations (Fransen & Edmonds, 1999).

(2) Uptake of exercise and PA over the last 12 months

In those with any self-reported joint pain exercise and PA as a treatment for their joint pain was also measured and used in the thesis as an indication of uptake of PA as a recommended treatment by the NICE guidelines (NICE, 2014). The uptake of PA as a treatment for joint pain was measured by asking responders to report whether they had received muscle strengthening exercises or general fitness exercises in the last 12 months for their joint pain or problems. In the thesis, responders were categorised as either receiving PA treatment or not receiving PA treatments.

(3) Global perceived change in joint problems

A global perceived change in joint problems over the last three months was a 1-6 single item scale asking participants to compare their change in joint problems from when they first responded to the baseline questionnaire. The scale ranged from a score of 1 to indicate a complete recovery to 6 indicating much worse joint problems. This gives a single numerical rating of change in participants during the trial of the study, with classification as those who have improved or those who have remained stable or worsened compared to entry into the trial.

(4) The IPAQ-SF

In the MOSAICS baseline and three month consultation survey the IPAQ-short form was used to measure levels of PA. The IPAQ-SF was developed as a generic patient reported outcome to measure levels of PA in international populations (Craig et al., 2003). The IPAQ-SF measures energy expenditure per week and can give a categorical score rating of an individual's level of weekly PA. The IPAQ-SF

contains four items which assess PA in sedentary activities, walking activities, moderate intensity activities and vigorous intensity activities, respectively.

(5) The PASE

The PASE is designed to be used specifically in adults aged 65 and over (Washburn et al., 1993). The PASE gives an output score of activity counts and scores usually range from 0-400 (Washburn et al., 1993). The PASE contains 32 items on PA in leisure, occupational and household activities.

The IPAQ and PASE are described and critically evaluated in the systematic review of this thesis (Chapter eight).

6.5 Summary

This chapter has outlined the protocol of the MOSAICS population survey that was used in aim one of this thesis, it has also described the protocol of the baseline and three month MOSAICS consultation questionnaire used in aim three. The main features of this chapter were:

- The methodology used in the MOSAICS research project that were used in this thesis
- A description of the self-reported outcome measurements used in the MOSAICS population survey and used in this thesis
- A description of the self-reported outcome measurements used in the MOSAICS consultation questionnaire that were used in this thesis.

The next three chapters describe the work conducted to address the three aims. These chapters use the methods described in Chapter five together with chapters seven and nine, use the data taken from the MOSAICS population survey and consultation questionnaires discussed in this chapter.

Chapter Seven: Describing levels of physical activity in community dwelling UK adults with and without self-reported joint pain

7.1 Introduction

It is unclear if joint pain symptoms reduce levels of PA and if more severe joint pain reduces levels of PA further. A number of issues from the literature have emerged, adults with OA in clinical settings would seem to participate in less PA (de Groot et al., 2008; Holsgaard-Larsen & Roos, 2012), although not all studies have found this (White et al., 2013). One study found high levels of PA in a hip or knee OA population in a hospital setting (Holsgaard-Larsen & Roos, 2012) compared to general population UK adult populations (BHF, 2012) and other studies in adults with clinical OA (de Groot et al., 2008; White et al., 2014). Previous OA studies measuring levels of PA have only focussed on adult populations with OA in the knee or hip OA, despite OA also affecting other joint sites (Thomas et al., 2004), particularly in the four sites focussed on by NICE in the OA treatment guidelines (NICE, 2014). Currently there have been no studies on the level of PA in the target population at a population survey level in the UK. It is not currently clear if joint pain in more areas compared to joint pain in the lower limb only (joint pain in feet, knees or hips), is associated with poorer PA levels, as this could signal more adults at high risk of low level of PA. It has previously been shown that more painful sites result in worse health outcomes (Kamaleri et al., 2008). As no studies have specifically focussed on describing the levels of PA in the target population at a population level rather than clinical, this chapter presents a description of the levels of PA in the target population.

7.2 Aim

To describe the self-reported levels of PA of UK community dwelling adults aged 45 years and over, with and without joint pain.

7.3 Objectives

To achieve this aim, three objectives were included:

1a. To describe the overall levels of PA in an adult population aged 45 years and over, with and without self-reported joint pain.

1b. To describe the physical and mental health status of adults aged 45 years and over with self-reported joint pain reporting different levels of PA.

1c. To describe and compare levels of PA in two subgroups of adults aged 45 years and over with self-reported joint pain: in the lower limb only and in generalised joint pain (upper and lower limb).

7.4 Statistical analysis

7.4.1 Study participants' inclusion and exclusion criteria

The inclusion criteria for this study were the same as for the MOSAICS project plus those that had responded to the MOSAICS population survey and completed the STAR questionnaire. Exclusion criteria for this study were: respondents who did not complete the STAR or respondents with completely missing data for joint pain and joint pain intensity questions.

7.4.2 Participant categories

In objectives 1a and 1b, respondents were separated into two groups: respondents who reported joint pain in the hip, knee, foot or hand and respondents who reported no joint pain in these sites. In objective 1c, those with self-reported joint pain were then subsequently separated into two further groups: those with joint pain in the lower limbs only (hip, knee or foot) and those with generalised reported joint pain in the upper and lower limbs (hip, knee or foot and hand).

7.4.3 Descriptive statistics

For descriptive statistics with continuous data, mean and standard deviations were calculated for normally distributed data. Median and interquartile ranges were provided in data that were non-normally distributed. Comparisons of two populations with continuous data were analysed using independent sample *t*-tests and the Wilcoxon rank-sum test. In categorical data the number and the percentage of responders within each category was displayed. To test associations between three categories of participants in categorical data, analyses were conducted using the chi-squared test.

7.4.4 Ordinal regression

For objective 1a, ordinal regressions were conducted for comparisons of levels of PA between respondents that had self-reported joint pain and those who did not report any joint pain. For objective 1c, ordinal regression models conducted to compare levels of PA between respondents with self-reported lower limb joint pain only and generalised joint pain compared to respondents with no reported joint pain as the reference group. Ordinal regressions were presented using odds

ratios (OR) of groups being less likely to be in the higher PA levels categories of the STAR questionnaire compared to the reference group. Four assumptions of the data for an ordinal regression were checked prior to ordinal regressions being conducted (Field, 2012). These assumptions were: (1) the dependent variable was measured to an ordinal level; (2) the independent variables were continuous, ordinal or categorical level data; (3) there was no multicollinearity between independent variables; and (4) there were proportional odds between groups, the independent variables having proportionally the same odds at each split between the different levels of the dependent variable.

A hypothetical example of proportional odds can be described for PA levels using the STAR instrument. The STAR has three ordinal levels, with levels 1 and 3 being the lowest and highest PA levels, respectively. Proportional odds assumes that the odds ratio for being in level 3 compared to being in levels 1 or 2 is the same as being in levels 2 or 3 compared to level 1. For example, in the STAR questionnaire proportional odds assumes that the odds ratio for being in the highest level of PA compared to being in the medium or lowest levels is the same as being in the highest or medium levels of PA compared to the lowest level. Assumption (1) has been met, as the classification of the STAR provides ordinal level data. Assumption (2) has also been satisfied, as the independent variables are either continuous, ordinal or categorical level data. The results and justification for using an ordinal regression for assumptions (3) and (4) are given in the results section of this chapter, page 124 and 130.

A number of confounding factors identified as affecting PA levels were controlled for in the ordinal regression model. The confounding factors for PA included in the

adjusted ordinal regression models were identified from previous studies, and included: age, with adults shown to be less active as age increases beyond 20 years of age (Haskell et al., 2007); gender, with female adults being less active compared to males (Ford et al., 2005); socio-economic status was controlled for as those with lower socio-economic status demonstrating lower PA levels (Ford et al., 1991); finally, BMI was also controlled, with being overweight or obese found to be associated with reduced levels of PA (Stamatakis et al., 2007).

7.4.5 Analysis of Variance

For objective 1b, a one-way analysis of variance (one-way ANOVA) was conducted in those with self-reported joint pain comparing the mean PCS and MCS of the SF-12 those with self-reported joint pain in the different PA categories of the STAR questionnaire. A Scheffe post hoc test (Field, 2012) was conducted to identify differences in PCS and MCS between the three individual categories of the STAR. An adjusted ANOVA was conducted for the same identified confounding factors as used in the adjusted ordinal regression model.

7.5 Results

7.5.1 Descriptive statistics of responders

The population survey was mailed out to 28,443 adults aged 45 and over registered at 8 general practices. There were 16,239 (57.1%) responders to the population survey. After the cleaning data of the responders there were 15,083 with complete data of the joint pain questions (92.9% of responders). Due to the STAR questionnaire being the instrument used for this study; those with missing

data in both of these variables were further excluded. There were 663 (4.4%) responders with missing STAR data and 208 (2.3%) responders with missing data for joint pain in the last 12 months in the hip, knee, hand or foot. As a result, there were 14,212 (94.2%) responders included in this study. Figure 7.1 displays the response rate of the MOSAICS population survey and table 7.1 displays the age and gender of responders and non-responders for the MOSAICS population survey.

Figure 7.1 Flowchart of MOSAICS population survey

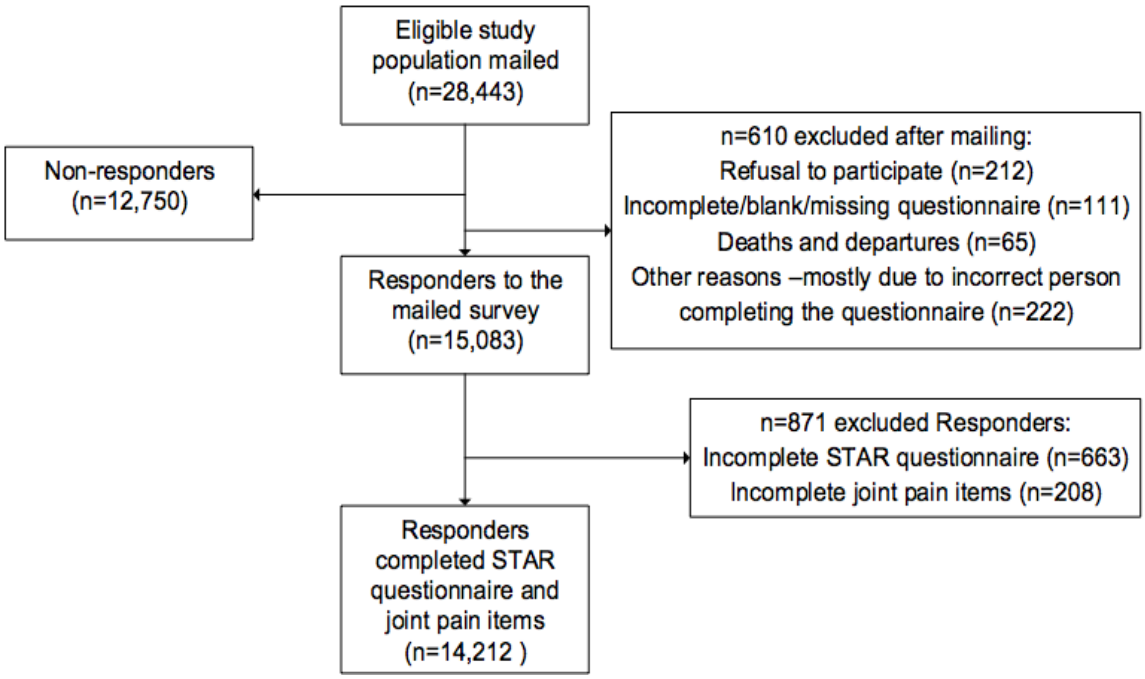


Table 7.1 Age and gender of responders and non-responders of the MOSAICS population survey

	Responder (n=14212, 50%)	Non-responders (n=14231, 50%)	P-value
Gender, n (%): Males Females	 6516 (45.8%) 7696 (54.2%)	 7396 (52.0%) 6835 (48.0%)	 <0.001*
Mean age, years (sd):	63.61 (11.15)	59.33 (11.50)	<0.001††
Age, years, n (%): 45-54 55-64 65-74 ≥75	 3462 (24.4%) 4355 (30.6%) 3793 (26.7%) 2602 (18.3%)	 6065 (42.6%) 4163 (29.3%) 2238 (15.7%) 1765 (12.4%)	 <0.001*

Key: *statistically significant below (p=<0.05) using chi-squared analysis. ††statistically significant below (p=<0.05) using independent *t*-test analysis. sd equals standard deviation.

Table 7.1 demonstrates that a statistically significant higher percentage of males and a lower percentage of females did not respond the MOSAICS population survey compared to those who did respond. Those who did not respond to the population survey were also significantly younger in age compared those who responded.

7.5.2 Objective 1a

Objective 1a described the overall levels of PA in an adult population aged 45 years and over with and without self-reported joint pain. Descriptive statistics of responders with and without reported joint pain are presented in Table 7.2. Table 7.3 displays further characteristics of respondents with reported join pain. Figure 7.2 displays the levels of PA in adults with and without self-report joint pain.

Table 7.2 Descriptive statistics of respondents in the MOSAICS population survey

	No reported joint pain in last 12 months	Reported joint pain in last 12 months	P-value
All responders, n (%)	2902 (20.4%)	11310 (79.6%)	
Gender, n (%): Males Females	 1492 (51.4%) 1410 (48.6%)	 5024 (44.4%) 6286 (55.6%)	 <0.001*
Mean age, years (sd):	62.20 (11.16)	63.97 (11.12)	<0.001ǂǂ
Age range, years:	45-100	45-96	
Age, years, n (%): 45-54 55-64 65-74 ≥75	 859 (29.6%) 877 (30.2%) 699 (24.1%) 467 (16.1%)	 2603 (23.0%) 3478 (30.8%) 3094 (27.4%) 2135 (18.8%)	 <0.001*
Mean BMI, kg/m^2 (sd):	25.64 (4.03) (n=2807)	27.23 (4.85) (n=10910)	<0.001ǂǂ
BMI, n (%): Underweight (<18.5kg/m^2) Healthy range (18.6-24.9kg/m^2) Overweight (25-29.9kg/m^2) Obese (≥30kg/m^2) Missing	 44 (1.5%) 1325 (45.7%) 1068 (36.8%) 370 (12.7%) 95 (3.3%)	 104 (1.0%) 3778 (33.4%) 40.4 (39%) 2615 (23.1%) 400 (3.5%)	 <0.001*
Mean IMD deprivation score (sd):	20811.16 (8020.43)	20183.12 (8311.19)	<0.001ǂǂ
IMD deprivation score Ranges:	237-32468	444-32468	
Mean health status (sd): Physical component score Mental component score	 52.30 (8.61) 52.74 (8.71) (n=2839)	 43.67 (12.41) 49.55 (10.55) (n=11058)	 <0.001ǂǂ <0.001ǂǂ
Level of physical activity, n (%): Inactive Somewhat Active	 233 (8.1%) 1075 (37.0%) 1594 (54.9%)	 1329 (11.8%) 5011 (44.3%) 4970 (43.9%)	 <0.001*

Key: *statistically significant below (p=<0.05) using chi-squared analysis. ǂǂstatistically significant below (p=<0.05) using independent *t*-test analysis. sd equals standard deviation.

There were a higher percentage of females self-reporting joint pain compared to those reporting no joint pain (55.6% compared to 48.6%). The mean age for those who reported no joint pain was significantly lower (*t*=7.67, df=14210, p=<0.001).

Mean BMI had 495 (3.5%) responders with missing data due to either height or weight not reported. Those with self-reported joint pain had also scored significantly higher in BMI compared to those with no reported joint pain (*t*=17.79, df=5105.42, p=<0.001). Both the PCS and MCS were significantly lower in those with self-reported joint pain (PCS; *t*=43.12, df=6219.50, p=<0.001, MCS; *t*=16.65, df=5186.61, p=<0.001). Finally, PA levels were compared between those with self-reported joint pain and those with no joint pain. Table 7.2 suggests those with joint pain are less likely to be active or meeting recommended levels of PA (Bull & Expert working panel, 2010) compared to those who reported no joint pain.

The location, intensity and distribution of pain intensity at individual joint sites in the hip, knee, hand or foot are displayed in Table 7.3.

Table 7.3 The location, intensity and distribution of joint pain in the last 12 months.

	Any joint pain in last 12 months (n=11310)
Site of joint pain, n (%):	
Hip	5682 (50.2%)
Knee	7757 (68.6%)
Hand	6209 (54.9%)
Foot	5773 (51.0%)
Number of joint pain sites, n (%):	
1 site	3522 (31.1%)
2 or more sites	7788 (68.9%)
Mean pain intensity in last month (0-10) (sd): (n=11241)	4.94 (2.70)
Pain intensity category in last month, n (%):	
0 No pain	273 (2.4%)
1-4 mild pain	5029 (44.5%)
5-10 moderate/ severe pain	5939 (52.5%)
Missing	69 (0.6%)
Joint pain category, n (%):	
Lower limb pain only	5101 (45.1%)
Generalised joint pain	5458 (48.3%)
Hand pain alone	751 (6.6%)

Key: sd equals standard deviation. In categorical variables for site of joint pain percentages are shown for each site in the total sample of those with any reported joint pain. In categorical variables for number of joint pain sites and joint pain category in last month percentages are shown down the column.

Table 7.3 demonstrates that in those with joint pain, the knee was the most common site of pain (68.6%) followed by the hand (54.9%), foot (51.0%) and then the hip (50.2%). A large percentage of those with joint pain (68.9%) reported having it in more than one site over the last 12 months. Mean (SD) pain intensity was 4.92±2.71 in those with reported joint pain. In those who had reported pain in the last 12 months, 238 (1.6%) had a pain intensity of 0 (no pain) over the last month; 40.3% of those with joint pain had their highest pain intensity classified as mild pain and 52.5% as moderate or severe pain intensity severity using Zelman et al. method (2005). Table 7.3 shows that 45.1% of those with reported joint pain

had lower limb pain only, 48.3% had generalised joint pain and only 6.6% reported joint pain in the hand alone.

Figure 7.2 Levels of PA in responders with and without self-reported joint pain

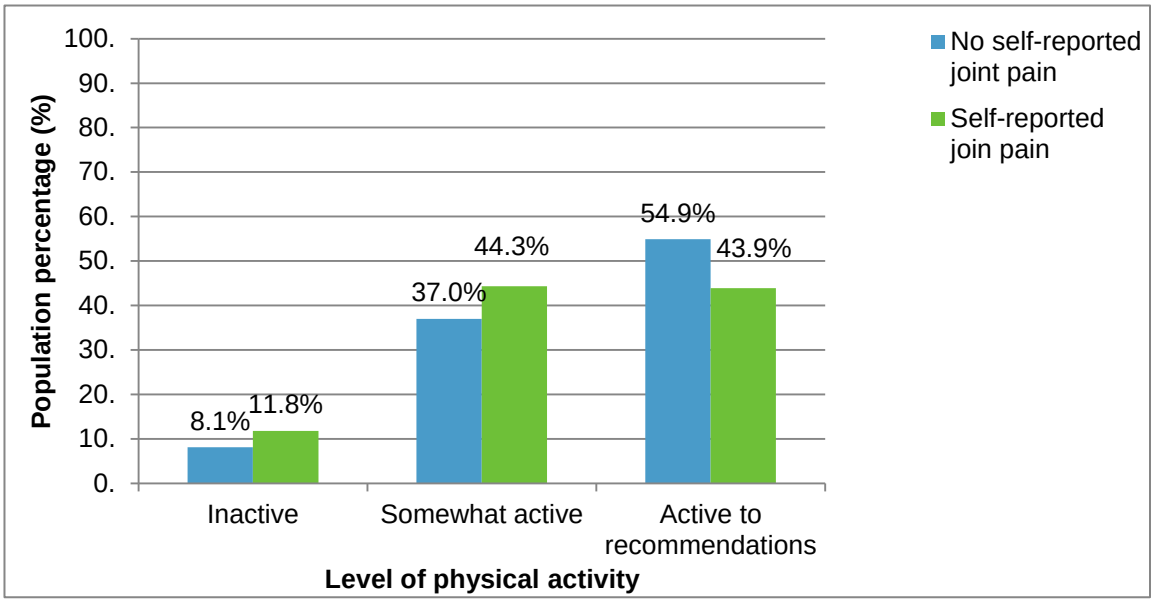


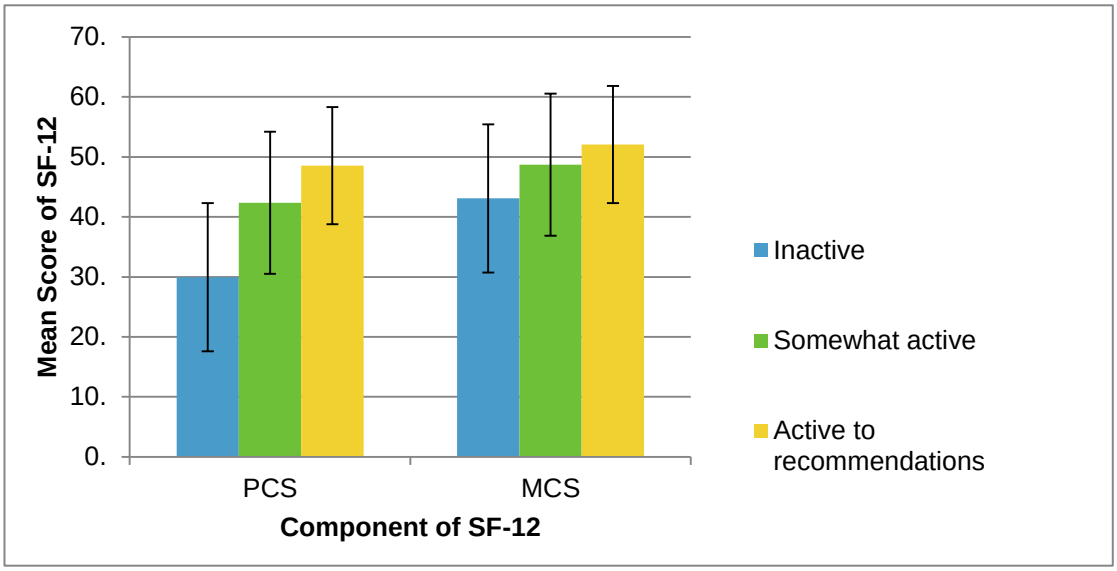
Figure 7.2 shows that higher proportions of responders with joint pain were inactive or somewhat active, according to STAR categories, compared to responders who did not report any joint pain. Lower proportions of responders with joint pain were achieving recommended levels of PA compared to those who did not report joint pain. Assumptions (3) and (4) were tested prior to the ordinal regression analyses being conducted. In the adjusted ordinal regression, assumption (3) was tested for collinearity between independent variables, using scatter graphs to assess relationships. There were no strong linear relationships (r=>0.7) identified between all of the independent variables. Assumption (4) was tested in the data using the test of parallel lines. For the ordinal regression, the test of parallel lines was not significant (x2=0.07, df=1, p=0.79), suggesting the proportional odds between joint pain groups were the same.

The crude ordinal regression demonstrated that responders with self-reported joint pain were less likely to be scored in the “somewhat active” or “active to recommendations” categories of the STAR, compared to participants with no reported joint pain (OR= 0.65, 0.60-0.70 95%CI). The adjusted ordinal regression showed respondents with self-reported joint pain were less likely to be scored in the “somewhat active” or “active to recommendations” categories of the STAR, compared with participants with no reported joint pain (OR= 0.78, 0.72-0.85).

7.5.3 Objective 1b

Objective 1b was to describe the physical and mental health status of adults aged 45 years and over with self-reported joint pain reporting different levels of PA. A one-way ANOVA was conducted in those with self-reported joint pain to compare PCS and MCS scores in the three PA categories using the STAR questionnaire. A Scheffe post hoc test was conducted to compare the individual difference between the three STAR categories. Figure 7.3 displays the mean MCS and PCS of the SF-12 in each of STAR levels of PA.

Figure 7.3 Mean PCS and MCS in different STAR categories in those reporting joint pain.



Key: error bars display standard deviations.

Those with joint pain who reported being inactive had lower mean PCS (29.93±12.10) compared to those who reported being somewhat active (42.36±11.85). Those reporting being inactive and somewhat active had lower PCS scores compared to those who had reported being active to the recommendations (48.55±9.76). The one-way ANOVA found the difference between groups in PCS to be statistically significant (f=1149.10, df=2, p=<0.001). The Scheffe post hoc analysis also found significant differences between each of the individual groups (<0.001). An adjusted one-way ANOVA for identified confounders also showed a significant difference in mean PCS between PA groups (f=1011.87, df=2, p=<0.001).

Those with reported joint pain who reported being inactive also reported in line with the findings of the PCS a lower mean score of the MCS (43.08±12.35) compared to those who reported being somewhat active (48.69±10.57). Both

those who reported being inactive and somewhat active had lower mean MCS compared to those reporting being active to recommendations (52.09±9.06). These differences in mean score across all the different categories of PA was found to be significantly different in a one-way ANOVA (*f*=427.20, df=2, p=<0.001). The Scheffe post hoc analysis also found significant differences between each of the individual groups (p<0.001). An adjusted one-way ANOVA for identified confounders also showed a significant difference in mean MCS between PA groups (f=458.82, df=2, p=<0.001).

7.5.4 Objective 1c

Objective 1c was to describe and compare levels of PA in two subgroups of adults aged 45 years and over with self-reported joint pain in the lower limbs only and generalised joint pain. Levels of PA were compared between those with lower limb joint pain only and those with generalised joint pain to those with no joint pain. Descriptive statistics for responders with lower limb joint pain only and generalised joint pain are displayed below in Table 7.4.

Table 7.4 Descriptive statistics for responders with lower limb pain only and responders with generalised joint pain.

	No reported joint pain, 2902 (20.4%)	Lower limb joint pain only, n=5101 (48.3%)	Generalised joint pain, n=5458 (51.7%)	P- value
Gender, n (%):				
Males	1492 (51.4%)	2540 (49.8%)	2142 (39.2%)	
Females	1410 (48.6%)	2561 (50.2%)	3316 (60.8%)	<0.001*
Mean age, years (sd):	62.20 (11.16)	63.05 (11.81)	64.92 (11.05)	<0.001ǂǂ
Age range, years:	45-100	45-96	45-100	
Age, years, n (%):				
45-54	859 (29.6%)	1322 (25.9%)	1095 (20.1%)	
55-64	877 (30.2%)	1604 (31.4%)	1651 (30.2%)	
65-74	699 (24.1%)	1300 (25.5%)	1572 (28.8%)	
75+	467 (16.1%)	875 (17.2%)	1140 (20.9%)	<0.001*
Mean BMI, kg/m^2 (sd):	25.64 (4.03) (n=2807)	27.17 (4.60) (n=4931)	27.51 (5.14) (n=5244)	0.001*
BMI, n (%):				
Underweight (<18.5kg/m^2)	44 (1.5%)	42 (0.5%)	54 (1.0%)	
Healthy range (18.6-24.9kg/m^2)	1325 (45.7%)	1667 (33.4%)	1754 (32.1%)	
Overweight (25-29.9kg/m^2)	1068 (36.8%)	2082 (40.5%)	2045 (37.5%)	
Obese (≥30-34.9kg/m^2)	370 (12.7%)	1140 (22.3%)	1391 (25.5%)	
Missing	95 (3.3%)	170 (3.3%)	214 (3.9%)	<0.001*
Mean IMD deprivation score (sd):	20811.16 (8020.43)	20608.46 (8212.82)	19605.97 (8424.81)	<0.001ǂǂ
Mean health status (sd):				
PCS	52.30 (8.61)	46.09 (11.55)	40.48 (12.97)	<0.001ǂǂ
MCS	52.74 (8.71) (n=2839)	50.88 (9.95) (n=4999)	48.03 (11.08) (n=5321)	<0.001ǂǂ
Mean pain intensity in last month (0-10) (sd):		4.31 (2.58)	5.78 (2.60)	<0.001ǂǂ
Pain intensity severity in last month, n (%):				
0 No pain		194 (3.8%)	33 (0.6%)	
1-4 mild pain		2718 (53.7%)	1790 (32.9%)	
5-10 moderate/ severe pain		2154 (42.5%)	3610 (66.5%)	<0.001*
Level of PA, n (%):				
Inactive	233 (8.1%)	501 (9.8%)	785 (14.4%)	
Somewhat	1075 (37.0%)	2245 (44.0%)	2451 (44.9%)	
Active	1594 (54.9%)	2355 (46.2%)	2222 (40.7%)	<0.001*

Key: sd equals standard deviation.*statistically significant below (p=<0.05) using chi-squared analysis. ǂǂstatistically significant below (p=<0.05) using independent *t*-test. In categorical variables for gender, age, BMI, pain intensity severity in last month and level of PA percentages are displayed down the columns. PCS =physical component score, MCS= mental component score.

Signficances testing performed in lower limb joint pain and generalised joint pain only, and did not include those with no reported joint pain.

Those with reported lower limb joint pain only and those with generalised joint pain were compared as shown in table 7.4, a higher percentage of females were categorised as generalised joint pain. Those with generalised joint pain were slightly but significantly older with a higher mean BMI. Those with generalised joint pain were also living in lower socio-economic areas, reported lower PCS and MCS in the SF-12, and reported higher joint pain intensity. Figure 7.4 displays the level of PA in adults with generalised joint pain lower, and a lower percentage active to recommended levels of PA compared to those with lower limb joint pain only. A higher percentage of adults with generalised joint pain reported being inactive compared to those with lower limb joint pain. Figure 7.4 displays the level of PA reported by responders with no reported joint pain, those with lower limb only and those with generalised joint pain.

Figure 7.4 Levels of PA in responders without self-reported joint pain and those with lower limb joint pain and generalised joint pain.

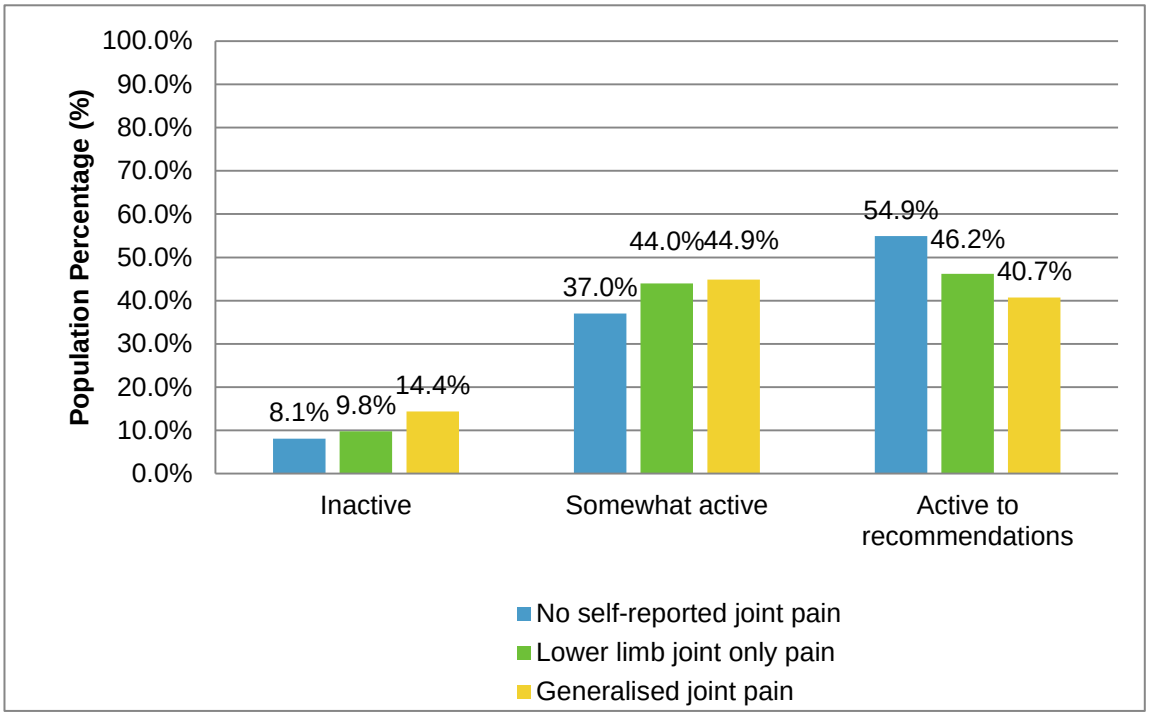


Figure 7.4 shows that those with generalised joint pain and lower limb only joint pain had higher percentages reporting as inactive and somewhat active compared to those who did not report joint pain.

Assumption (4) was tested for the ordinal regression using the test of parallel lines, however a significant difference was found (x2=12.10, df=1, p=0.001), suggesting the proportional odds between joint pain groups differed. Because the assumption of proportional odds was not met in this model, the STAR data were dichotomised into “active to recommendations” and “below activity levels of recommendations”; this was conducted by collapsing the “inactive” and “somewhat active” categories into one group. A binary logistic model was then conducted to evaluate the odds of those with generalised joint pain being less likely to be responsive to PA recommendations. Adults with generalised joint pain were found to be significantly

less likely to achieve recommended levels of PA, compared to those with lower limb joint pain only (OR= 0.80, 0.74-0.87 95%CI). An adjusted binary logistic model also showed a significant difference (OR= 0.89, 0.82-0.97 95%CI).

7.6 Discussion

This section discusses the main findings of this study in the overall aim of this chapter's study as outlined in section 7.2 and the context of each of the objectives outlined in section 7.3. This section then makes comparisons of the findings of this study in relation to other relevant research. The strengths and limitations together with the possible future research based on this study are discussed in Chapter ten of the thesis.

7.6.1 Main findings

Prior to this study the levels of PA in the target population had not been reported in the UK. Previous studies had reported on PA levels of adults with lower limb OA (De Groot et al., 2005; Rosemann et al., 2007; Dunlop et al., 2011a; Holsgaard-Larsen & Roos, 2012; White et al., 2013). These studies' participants were in clinical settings rather than at a population level and focused on OA in the knee and hip joint, this study has focused on the four joint pain sites NICE used in their latest OA management guidelines (NICE, 2014). It is not clear whether joint pain is associated with lower levels of PA compared to those with no joint pain. If joint pain is associated with lower levels of PA, then individuals are also at risk of poorer health associated with joint pain and lower PA levels. In their latest guidelines on OA management, NICE recommended future research focussing on multisite OA and joint pain (NICE, 2014). Prior to this study it was unknown if

adults aged 45 years and over with multisite joint pain in upper and lower limbs would be associated with lower levels of PA compared to those with lower limb joint pain only. Previous meta-analysis had shown that PA interventions improve disability and pain intensity in lower limb OA (Uthman et al., 2013) and Dunlop et al. (2011a) had shown that physical functioning measured using gait speed is higher in adults aged 45 years and over with OA that report higher PA levels. Prior to this study it was not clear at a population level if PA levels are associated with physical health-related quality of life (QoL).

This study has shown that, at a population level, adults aged 45 years and over with self-reported joint pain are more likely to report lower PA levels compared to those with no reported joint pain. This is important as adults aged 45 years and over with joint pain are a population that would gain greater benefits from being physically active in terms of reducing pain and disability. Using the STAR questionnaire to measure PA levels only 11.8% of adults aged 45 years and over with joint pain reported low levels of PA, although this was a higher percentage compared to respondents with no reported joint pain. The benefit of being more physically active for adults aged 45 years and over with joint pain was shown in this study. Adults aged 45 years and over with reported joint pain who reported the higher levels PA in the STAR also reported higher physical and mental health status using the SF-12. This shows the importance of PA in adults aged 45 years and over with joint pain at a population level. It is also interesting to note that those who reported being active to recommendations had SF-12 PCS and MCS close to 50, which represents the average values of general adult populations aged 18 and over. This suggests that higher levels of PA are associated with improving physical and mental health status in adults aged 45 years and over with joint pain to that of

adults without joint pain. When compared to respondents with lower limb joint pain only, those with generalised joint pain appeared to be a more severe group in terms of joint pain with physical and mental health status. Those with generalised joint pain also reported lower PA levels compared to those with lower limb joint pain only. This suggests that not only are those with joint pain more likely to be lower in PA levels but those with more severe joint pain in more sites are at higher risks of inactivity or low PA levels. Given that those with more severe joint pain have higher intensity of joint pain and lower physical and mental health status, this is a missed opportunity as increasing their PA would help in reducing their joint pain and disability, as well as reducing risk to other long term conditions that were discussed in Chapter two.

7.6.2 Comparison of findings with other studies

When comparing the findings of this study with other previous research, comparisons were made with the findings of each objective in this study.

(1) Objective 1a

The prevalence of joint pain in the hip, knee, hand or foot in an adult population for this study was 79.6%. This is a higher percentage compared to a US study that had found the prevalence of OA affecting 33.6% adults aged 65 years and over (Lawrence et al., 2008). Data in those with joint pain displayed suggest that those who reported joint pain in the hip, knee, hand or foot are in keeping with a population with OA (Neogi & Zhang, 2013). The population survey categorised respondents as to whether individuals had reported any joint pain in the past 12 months. No data on the conditions or causes of the joint pain were recorded.

Although OA is the most common cause of musculoskeletal pain in adults aged 45 years and over, some respondents could have had other musculoskeletal conditions that caused their joint pain. Many of these other musculoskeletal conditions may co-exist with OA.

The joint pain prevalence in this study is also higher compared to UK pain prevalence of 66.2% in a population survey study (Thomas et al., 2004). Possible explanations for this high prevalence of joint pain reported in this study could be the recall period of any pain around the joint over 12 months; this can include conditions such as OA, polymyalgia rheumatica, gout, fibromyalgia, carpal tunnel, ankylosing spondylitis and other pain around the joints. Joint pain may have also been caused by injury to the joint; in adults aged 45 years and over injury to joints can be caused by falls. No data can be found on prevalence of injury to the hip, knee, hand or foot in adults aged 45 years and over. Falls in adults aged 65 years and over are estimated at 40% (Rubenstein, 2006). Non-responder bias may have also interacted with this high prevalence with the distribution of non-responders and presences of joint pain unknown.

A number of other musculoskeletal conditions can also affect joint pain in adults aged 45 and over, these conditions are not as high in prevalence of OA in adults aged 45 and over, so it is likely that OA was the most prevalent condition of the population survey respondents with joint pain. Other musculoskeletal conditions that could have contributed to joint pain in respondents with reported joint pain include polymyalgia rheumatica which can affect the hip, knee, hand and foot in less than 1% of adults aged 50 years and over; gout affecting less than 2.5% of adults aged 18 years with prevalence increasing with age (Kuo et al., 2013);

fibromyalgia affects 3.3% of adults when adjusted for gender and age and carpal tunnel syndrome is shown to affect 1.5% of adults (Lawrence et al., 2008). Ankylosing spondylitis can also cause knee pain to less than 1% of adults (Kaipiainen-Seppanen et al., 1997).

In the BHF report on the self-reported PA levels in UK adults, 41% of males and 32% of females aged 45-55 years old were achieving recommended levels of PA (Bull & expert working group, 2012), with the number achieving recommendations lowering with age (BHF, 2012). This represents a lower level of PA compared to the MOSAICS study population. A UK based RCT measured self-reported levels of PA found 28% of adults with a mean age of 54 years achieved recommended levels of PA (Bull & expert working group, 2012). Furthermore, 27% were moderately active or inactive below the recommendations and 45% were totally inactive using the General Practice PA Questionnaire (GPPAQ) (Bull & Milton, 2010). These studies show a lower level of self-reported PA compared the MOSAICS population in groups with and without self-reported joint pain. Comparing findings in this thesis to PA data for adults based in the UK (Bull & Milton, 2010; BHF, 2012) suggest that levels of PA could be overestimated by the STAR. A different rationale could be that this population has higher PA levels compared to other adults populations in the UK or that the population overestimated their PA in the self-report survey. These alternative explanations would not seem to be plausible, because other characteristics, such as SF-12 scores, BMI and pain intensity seemed to be the similar to other UK studies in adult populations (Gandek et al., 1998; Thomas et al., 2004; Wang et al., 2011).

Helmerhorst et al. (2012) reported in a systematic review of self-report PA instruments for adults, that many instruments have weak validity compared to accelerometers, with self-report instruments overestimating levels of PA. This could be a possible explanation for the high self-report PA levels reported in this thesis.

The STAR questionnaire was developed and evaluated in an American adult population, was found to be reliable in test-retesting and valid compared to accelerometers (Matthews et al., 2005). Similar levels of PA were reported in adults with a mean age of 46 years to those in this thesis (Matthews et al., 2005).

(2) Objective 1b

Physical health and PA levels have previously been investigated using gait speed and self-report PA in US adults aged 45 years and over with knee OA (Dunlop et al., 2011). Gait speed was higher in those that reported higher PA levels, higher gait speed indicating higher physical functioning. In this study, respondents with joint pain who were active to recommendations (Bull & expert working panel, 2012) reported higher physical health scores on the SF-12. Rosemann et al. (2008) also reported this to be the case in German adults aged 45 years and over with lower limb OA. Higher health status and QoL using the Arthritis Impact Measurement Scale was associated with higher self-report PA levels. Higher PA levels have been associated with higher physical functioning in exercise interventions compared to controls with lower limb OA (Uthman et al., 2013).

(3) Objective 1c

In this thesis, of those who reported no joint pain only 8.0% were physically inactive, 37.0% were classified as somewhat active and 54.9% were classified as active, attaining recommendations for PA. Those with lower limb joint pain had higher odds of being inactive or somewhat active compared to those with no joint pain. Those with more severe generalised joint pain had higher odds of being inactive or somewhat active compared to those with no joint pain and those with lower limb joint pain only. When examining studies using objective measures of PA in those with hip or knee OA, the responders with joint pain, self-reported higher levels of PA compared to the objective measure PA in hip or knee OA populations (Dunlop et al., 2011a; White et al., 2013). The levels of PA reported by those with joint pain in this study were lower compared to a study in Sweden that used the Sensewear activity monitor (Holsgaard-Larsen & Roos, 2012). The Sensewear activity monitor was shown to be an overestimation of PA in adults aged 45 years and over with OA (Hermann et al., 2014). This suggests that the population with joint pain in this study had either higher levels of PA or had overestimated their self-reported PA levels compared to adults aged 45 years and over with OA.

The levels of PA in the joint pain populations in this thesis were more similar to those reported in a study using the IPAQ in adults with hip or knee OA (Rosemann et al., 2007). Rosemann et al. (2007) reported that 38% of individuals were categorised as active. When compared to the 54.9% categorised as active to recommendations (Bull & expert working group, 2012) in respondents with self-

reported joint pain, the PA levels in the MOSAICS population survey were still higher.

In contrast to the findings here, White et al. (2013) suggested that in US adults aged 45 years and over the presence of OA and the increasing intensity and severity of joint pain does not affect levels of PA. In contrast, De Groot et al. (2005) used objective measures to assess PA levels in adults aged 45 years and over from Holland and found that OA and joint pain was associated with reduced levels of PA. In this thesis adults aged 45 years and over with joint pain were found to have significantly lower levels of PA compared to adults with no reported joint pain. This was also the case when controlling for confounding factors: age, gender, BMI and socio-economic status.

Previous research has also shown that additional sites of pain are associated with worsening health status, less healthy lifestyle and disability (Kamaleri et al., 2008). The findings here are also supported by other work that found outcomes and healthy lifestyle are reduced with more sites of pain (Kamaleri et al., 2008).

7.7 Conclusion

This chapter has described the levels of adults aged 45 years and over with and without joint pain. Adults aged 45 years and over with joint pain are less likely to be more active compared to adults aged 45 years and over with no reported joint pain. In those with reported joint pain, lower levels of PA are associated with worse physical and mental outcomes. Those with generalised joint pain are less likely to be active compared to those with lower limb joint pain and those with no joint pain. Levels of PA in the population survey appeared to be overestimated

compared with other self-report measures in the UK target population. There is a need for an accurate measure of PA in adults aged 45 years and over with joint pain.

Chapter eight: Self-report measures of physical activity in adults with osteoarthritis and joint pain. A systematic review.

8.1 Introduction

This chapter discusses the methods, results and findings of a systematic review of self-report instruments of PA in the target population. In the analysis of the population survey in Chapter seven the STAR questionnaire was found to overestimate levels of PA in adults aged 45 years and over (with and without joint pain) compared to national data (BHF, 2012).

Systematic reviews have been published reporting the evidence for appropriate methods and instruments that can be used for assessing levels of PA in the young, adult and elderly populations (Forsen et al., 2010, Helmerhorst et al., 2012, Van Poppel et al., 2010, Chinapaw et al., 2010). All of these systematic reviews used a search strategy to identify measurement properties of self-report PA instruments in their populations of interest. General findings in these systematic reviews were that self-report PA instruments show good reliability in test-retest evaluations, particularly in self-report PA instruments for younger age groups aged 18 years or below, and adult populations aged 18-65 years, only moderate correlations were found in the self-report PA instruments with objective measurements (Chinapaw et al., 2010; Van Poppel et al., 2010; Helmerhorst et al., 2012). In elderly adult populations, although there were some studies on the measurement properties of self-report PA instruments, the number of high quality studies was lacking (Forsen et al., 2010).

A systematic review on the measurement of PA in populations with OA has been previously published (Terwee et al., 2011). The review gives an overview of different instruments with evidence of measurement properties of assessing PA in populations with OA of the hip and the knee. The review focused on evaluating the evidence of instruments' reliability and validity from previously published literature, responsiveness was not evaluated (Terwee et al., 2011). The review included instruments that were single-item instruments (scales), multi-item instruments (short questionnaires) and pedometers within a population with hip and knee OA. This allowed for decision making when identifying which instruments would be appropriate for an adult population with knee or hip OA based on the measurement properties of those instruments. The review concluded that there were not enough high quality studies evaluating measurement properties of PA instruments in OA populations (Terwee et al., 2011).

Terwee et al. (2011) only focused on hip and knee OA as a target population and did not include studies on generalised OA or in different sites of OA or peripheral joint pain in the hand or foot which are also common sites of OA (Arden & Nevitt, 2006). The PASE instrument which has been previously used in OA studies (Dunlop et al., 2011) was not captured in the review. Widening the inclusion criteria for studies on adults with OA and populations 45 years and over may have increased the amount of instruments and findings could be still be generalisable to adults with OA. Given the potential limitations of Terwee et al. (2011) review for populations with OA, a new systematic review was undertaken.

8.2 Aim

To evaluate the measurement properties of reproducible self-report PA instruments in adults aged 45 years and over with and without joint pain or OA.

8.3 Objectives

The following two objectives were addressed to achieve the aim:

2a. To identify self-report instruments using a systematic review of measures of PA previously used in adults aged 45 years and over with and without joint pain or OA.

2b. To describe the measurement properties of the self-report instruments identified by the systematic review measuring levels of PA in adults aged 45 years and over with and without joint pain or OA.

8.4 Methodology

To achieve the two objectives of this systematic review, a two stage search and review was conducted. Stage A identified all self-report instruments of PA used in published research in the area of the target population. Stage B of the systematic review then identified all the evidence on the measurement properties of those identified instruments. Selection criteria, search strategy, study selection, data extraction and quality assessment are described separately.

8.4.1 Stage A

(1) Stage A: Objective

The objective of Stage A was to identify self-report instruments using a systematic review of measures of PA previously used in adults aged 45 years and over with and without joint pain or OA.

(2) Stage A: Selection criteria

The selection criteria for the first stage of the systematic review is described in table 8.1 and included the relevant target population, outcomes of interest, study design and language. These criteria were agreed by consensus with the reviewing team and using Terwee et al.'s (2011) systematic review (See Table 8.1).

Table 8.1 Selection criteria for articles in Stage A.

Inclusion	Exclusion
Age range that includes participants 45 years old or over.	Over 50% of the study participants with inflammatory arthritis.
At least 50% of the study participants have OA or joint pain in the foot, knee, hip and hand.	A measure of physical fitness rather than a measure of daily PA participation.
Measurement instrument of PA using self-reported questionnaire.	Objective measures of PA. For example, accelerometers and calorimetry.
Self-reported PA used as a primary or secondary outcome measure.	In any other language than English.
All research settings (hospital, primary care, community settings, etc.)	Date range for articles will be from the databases' starting date until 18/12/2011, the date the search was conducted.
All quantitative research methodologies (RCTS, cross-sectional, etc.)	Case study research design of a single subject.

(3) Stage A: Inclusion criteria

The age 45 years or over represents the clinical definition of OA can be considered without x-ray (NICE, 2014). Studies with at least 50% of study participants defined with OA or joint pain were included. Self-reported PA instruments can be used in many different settings and different type of studies, so no limitations on study design or setting were used.

(4) Stage A: Exclusion criteria

Physical fitness can be described as the ability of an individual to perform daily activities and represents what an individual can achieve, representing a health status (Caspersen et al., 1985). PA is the musculoskeletal movement that increases energy expenditure over a period of time (Caspersen et al., 1985); within this review the focus was on PA only. Objective measure of PA, which include accelrometers and pedometers, were also excluded as many would not be of practical use within primary care clinical research.

The review was limited to articles written in the English language, although this introduces a language bias, the aim of the review was to find an appropriate measure of PA within a UK primary care setting, therefore, it was considered to be justifiable.

Articles were retrieved from the databases from date of origin to the end date of the search (18/12/2011) so that all articles could be identified.

(5) Stage A: Data Sources

The data sources for Stage A comprised of two separate searches: 1) electronic database searches and 2) hand-searching of references listed in selected articles. Databases were selected as being primary sources of articles for the review with the electronic databases used for stage A of this systematic review agreed by the reviewing team (see table 8.2).

Table 8.2 Electronic databases used for Stage A.

Database title	Acronym
Allied and Complementary Medicine	AMED
British Nursing Index	BNI
Cumulative Index of Nursing and Allied Health online	CINAHL
Cochrane Library, database of abstracts of reviews of effectiveness	Cochrane Library
Cochrane controlled trials register	Cochrane Register
Centre for Reviews and Dissemination reviews database	CRD
Execpta Medica online	Embase
The Health Management Information Consortium	HMIC
Index Medica online	Medline
Physiotherapy evidence database	PEDRO
American Psychological Association PsycINFO	PsycINFO
SPORTDiscus	Sportdiscus
Web of Science	WoS

(6) Stage A: Search strategy

A search strategy was constructed by identifying key criteria using the PICOS (Population, Intervention, Control/Comparison, Outcome and Setting) method (University of York Centre for Reviews, and Dissemination, 2009). PICOS allows for full identification of the required key terms based on the inclusion criteria. The PICOS method identified broad terms for Stage A search strategy which was agreed by the reviewing team and shown in table 8.3.

Table 8.3 PICOS for stage A.

PICOS	
Population:	Joint pain or OA in knee, hip, hand or foot
Intervention:	NA
Control/Comparison:	NA
Outcome:	Measurement of PA
Setting:	Any setting

NA=not applicable.

Previous systematic reviews within the Cochrane database were searched to identify established terms to locate articles for this systematic review. The previous systematic reviews that were used included: Fransen & McConnell (2008), Ortega et al. (2010) and Reichenbach et al. (2010). These terms were then pilot-tested for use with the databases MEDLINE, EMBASE and Web of Science to find the precision and sensitivity of the terms. Search terms that did not find any articles or brought a high number of inappropriate articles were taken out of the final search. The search terms were developed using the MEDLINE's medical subject headings (MESH), other search terms were established from MEDLINE's MESH for use in the other electronic databases. The final search terms were used and, if needed,

modified to fit the specific search setup across all of the databases used for the review. The search terms developed for the MEDLINE database can be seen in Box 8.1.

Box 8.1 Search terms for MEDLINE database (using OvidSP software)

1. exp osteoarthritis/
2. osteoarthr*.ti,ab
3. (degenerative adj2 arthritis).ti,ab
4. arthrosis.ti,ab
5. ((knee* or hip* or foot* or hand*) adj3 (pain* or ach* or discomfort*)).ti,ab
6. ((knee* or hip* or foot* or hand*) adj3 stiff*).ti,ab
7. “physical* activ*”.ti,ab
8. “physical therapy”.ti,ab
9. exercis*.ti,ab
10. rehabilitation.ti,ab
11. “leisure activ*”.ti,ab
12. “physical training”.ti,ab
13. 1 OR 2 OR 3 OR 4 OR 5 OR 6
14. 7 OR 8 OR 9 OR 10 OR 11 OR 12
15. 13 AND 14

Terms: exp – Explode of Mesh. ti – any word in title. ab – any word abstract. * - truncation symbol. sh – Mesh term.

Hand-searching was conducted on articles selected for full review after passing a study selection process at title, abstract and full article level. The reference lists of the articles were scanned and titles deemed relevant were included in the next step.

(7) Stage A: Study Selection

Study selection was done in three primary parts: title selection, abstract selection and full article selection. The aim for the study selection was to include all relevant studies in the review. Study selection was conducted in stage A by the primary reviewer (RS) and a secondary reviewing team of a second reviewer (KD), a third reviewer (MH) and a fourth reviewer (GH).

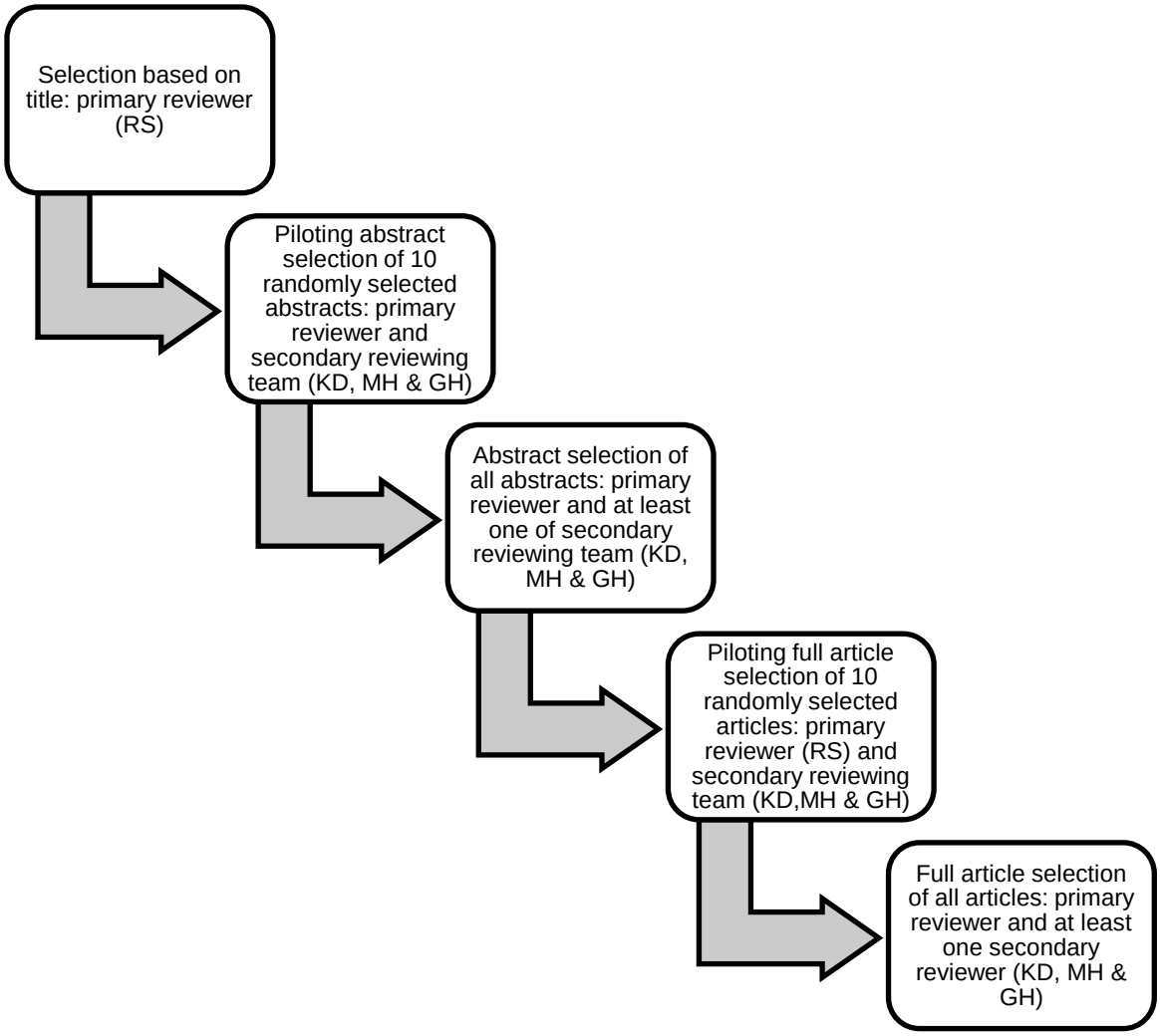
(8) Stage A: Selection based on titles

Title selection: the primary reviewer (RS) reviewed the titles identified by the search strategy. Titles of studies were considered for inclusion based on the eligibility. To reduce risk of error and bias, articles were only excluded if the title explicitly included something within the exclusion criteria or was far removed from any items in the inclusion criteria. The title selection process was conducted using References Management software, RefWorks (RefWorks 2.0, USA, 2010).

(9) Stage A: Article selection

Figure 8.1 displays a summary of the steps taken in the Stage A study selection process.

Figure 8.1 Flowchart of study selection process for Stage A.



Article selection was conducted using initial piloting prior to selection of all articles. In the article selection pilot, 10 full articles were randomly selected and assessed by all four reviewers using the inclusion and exclusion criteria and the same article selection form as used in the abstract selection process (Appendix 1.5). A consensus meeting between the primary reviewer, the secondary, the third and the fourth reviewer took place with the aim to achieve agreement in article selection.

Once agreement had been met, assessment of all articles was then conducted by the primary (RS), third reviewer (MH) and fourth reviewer (GH) independently. After the articles had been independently reviewed, the primary reviewer (RS) compiled the article selection forms and checked for cases of disagreement. Where there were articles with a disagreement, a consensus meeting took place. Articles where disagreement was found were discussed with each individual giving rationale for their decision and a discussion took place with the primary reviewer (RS), third reviewer (MH) and fourth reviewer (GH). The second reviewer (KD) assisted where agreement between the primary reviewer, third review and fouth reviewer could not be reached. Issues included clarification of inclusion and exclusion criteria, consistent application of criteria, matters of interpretation. In the pilot of article selection, all articles were discussed to identify both areas of agreement as well as disagreement.

(10) Stage A: Data extraction

The aim of Stage A in the systematic review was to identify all self-reported instruments of PA used in previous OA or joint pain research. A data extraction sheet was constructed specifically for this systematic review (Appendix 1.6). It recorded: the study setting, study type (RCTS, cross sectional studies, longitudinal studies, cohort studies or systematic reviews), which population was studied (age, diagnosis) and what self-reported measure(s) of PA were used in the study. In stage A of the systematic review no quality assessment of articles was required; the bias or methodological error within a study was not important to the outcome of the first stage, only the presence of a self-reported measure of PA was needed.

8.4.2 Stage B

(1) Stage B: Objective

From the self-report instruments identified in Stage A, an assessment of the measurement properties of the identified measures of PA was conducted in Stage B. The objective of Stage B was to describe the measurement properties of the self-report instruments identified by the systematic review measuring levels of PA in adults aged 45 years and over with and without joint pain or OA.

(2) Stage B: Selection criteria

Eligibility criteria were developed to identify the studies relevant to the systematic review's research question. The selection criteria for articles in Stage B were more detailed in comparison to Stage A to specifically identify articles evaluating measurement properties of self-reported PA questionnaires in the target populations (Table 8.4).

Table 8.4 Eligibility criteria for selection of articles on measurement properties of PA instruments in adults aged 45 years and over with and without joint pain or OA

Inclusion	Exclusion
Self-reported PA instruments identified in Stage A of the systematic review.	Non-English language articles.
The self-reported PA instruments in these studies should assess PA in at least one of the following: frequency of activity, duration, intensity of activity.	Objective measures of PA, PA diaries and unidentified items or questionnaires that could not be referenced
Studies that examine the instrument's qualitative attributes or measurement properties	Study participants defined by a health condition not joint pain or osteoarthritis, for example a study population with cardiovascular disease.
At least 50% of participants aged at least 45 years old or above. Or if these data are unavailable; the mean age of the population aged 45.0 years or above.	Commentary or letters to the editors.
Study participants with joint pain or osteoarthritis, or community dwelling adult populations.	Non-original research articles. (Reference lists were examined for relevant articles).

The review examined all measurement properties in studies of adult populations (45 years old and over). Adults aged 45 years and over was selected as from 45 years of age joint pain and OA becomes more prevalent (Arden & Nevitt, 2006) and NICE defines OA and joint pain occurring in this age group (NICE, 2014).

(3) Stage B: Search strategy

The search strategy for stage B of the review was constructed using a method designed and developed by Terwee et al. (2009) for high sensitivity in finding articles on measurement properties of measurement tools, such as PA. This systematic review combined the filter for measurement properties developed by

Terwee et al. (2009), along with the name of the instrument identified in Stage A of this review.

No construct search terms and population terms were used as the construct did not need defining because the instrument was defined instead. Filtering for the study participants took place at the article evaluation stage during title and abstract reviewing.

Electronic databases selected for Stage B included MEDLINE EMBASE, and Web of Science. No other electronic databases were included as a pilot of the search strategy identified no additional articles relevant to the review from other databases.

The search term strategy used for Stage B of the systematic review in MEDLINE is displayed below in Box 8.2; terms were modified in EMBASE and Web of Science to suit each database's medical subject headings.

Box 8.2 MEDLINE search terms for Stage Two.

1. Instrumentation.sh OR
2. Methods.sh OR
3. "Validation Studies".pt OR
4. "Comparative Study".pt OR
5. "psychometrics".sh OR
6. psychometr*.ti,ab OR
7. clinimetr*.tw OR
8. clinometr*.tw OR
9. "outcome assessment (health care)".sh OR
10. outcome assessment.ti,ab OR
11. outcome measure*.tw OR
12. "observer variation".sh OR
13. observer variation.ti,ab OR
14. "Health Status Indicators".sh OR
15. "reproducibility of results".sh OR
16. reproducib*.ti,ab OR
17. "discriminant analysis".sh OR
18. reliab*.ti,ab OR
19. unreliab*.ti,ab OR
20. valid*.ti,ab OR
21. coefficient.ti,ab OR
67. kappa.ti,ab OR
68. kappa's.ti,ab OR
69. kappas.ti,ab OR
70. repeatab*.ti,ab OR
71. ((replicab*.ti,ab OR repeated.ti,ab) AND (measure.ti,ab OR measures.ti,ab OR findings.ti,ab OR result.ti,ab OR results.ti,ab OR test.ti,ab OR tests.ti,ab)) OR
72. generaliza*.ti,ab OR
73. generalisa*.ti,ab OR
74. concordance.ti,ab OR
75. (intraclass.ti,ab AND correlation*.ti,ab) OR
76. discriminative.ti,ab OR
77. "known group".ti,ab OR
78. factor analysis.ti,ab OR
79. factor analyses.ti,ab OR
80. dimension*.ti,ab OR
81. subscale*.ti,ab OR

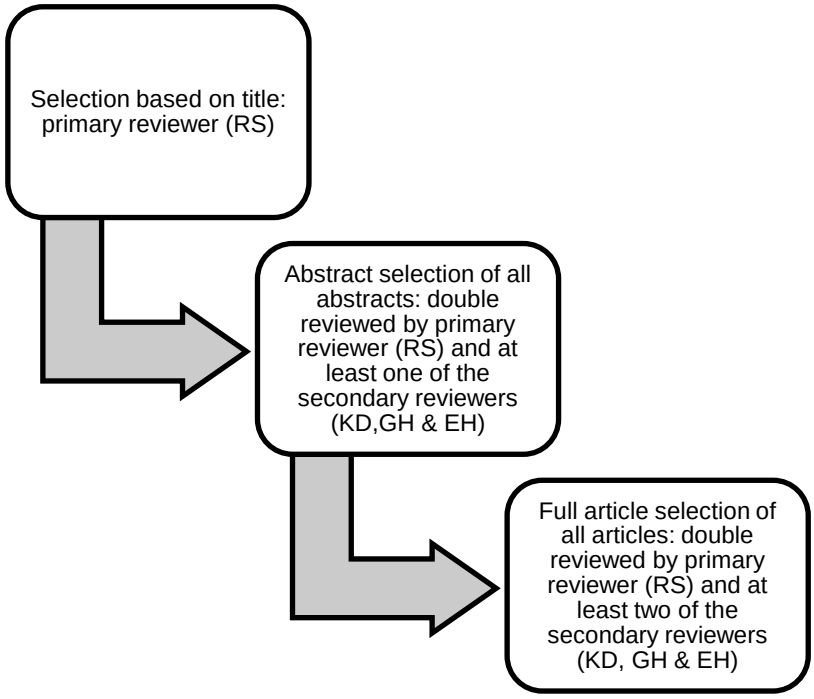
22. homogeneity.ti,ab OR
23. homogeneous.ti,ab OR
24. “internal consistency”.ti,ab OR
25. (cronbach*.ti,ab AND (alpha.ti,ab OR alphas.ti,ab)) OR
26. (item.ti,ab AND (correlation*.ti,ab OR selection*.ti,ab OR reduction*.ti,ab)) OR
27. Agreement.ti,ab OR
28. Precision.ti,ab OR
29. imprecision.ti,ab OR
30. “precise values”.ti,ab OR
31. test– retest.ti,ab OR
32. (test.ti,ab AND retest.ti,ab) OR
33. (reliab*.ti,ab AND (test.ti,ab OR retest.ti,ab)) OR
34. Stability.ti,ab OR
35. Interrater.ti,ab OR
36. inter-rater.ti,ab OR
37. intrarater.ti,ab OR
38. intra-rater.ti,ab OR
39. intertester.ti,ab OR
40. inter-tester.ti,ab OR
41. intratester.ti,ab OR
42. intra-tester.ti,ab OR
43. interobserver.ti,ab OR
44. inter-observer.ti,ab OR
45. intraobserver.ti,ab OR
46. intertechnician.ti,ab OR
47. inter-technician.ti,ab OR
48. intratechnician.ti,ab OR
49. intra-technician.ti,ab OR
50. interexaminer.ti,ab OR
51. inter-examiner.ti,ab OR
52. intraexaminer.ti,ab OR
53. intra-examiner.ti,ab OR
54. interassay.ti,ab OR
55. inter-assay.ti,ab OR
56. intraassay.ti,ab OR
57. intra-assay.ti,ab OR
58. interindividual.ti,ab OR
59. inter-individual.ti,ab OR
60. intraindividual.ti,ab OR
61. intra-individual.ti,ab OR
62. interparticipant.ti,ab OR
63. inter-participant.ti,ab OR
64. intraparticipant.ti,ab OR
65. intra-participant.ti,ab OR
82. (multitrait.ti,ab AND scaling.ti,ab AND (analysis.ti,ab OR analyses.ti,ab)) OR
83. item discriminant.ti,ab OR
84. interscale correlation*.ti,ab OR
85. error.ti,ab OR
86. errors.ti,ab OR
87. “individual variability”.ti,ab OR
88. (variability.ti,ab AND (analysis.ti,ab OR values.ti,ab)) OR
89. (uncertainty.ti,ab AND (measurement.ti,ab OR measuring.ti,ab)) OR
90. “standard error of measurement”.ti,ab OR
91. sensitiv*.ti,ab OR
92. responsive*.ti,ab OR
93. ((minimal.ti,ab OR minimally.ti,ab OR clinical.ti,ab OR clinically.ti,ab) AND (important.ti,ab OR significant.ti,ab OR detectable.ti,ab) AND (change.ti,ab OR difference.ti,ab)) OR
94. (small*.ti,ab AND (real.ti,ab OR detectable.ti,ab) AND (change.ti,ab OR difference.ti,ab)) OR
95. meaningful change.ti,ab OR
96. “ceiling effect”.ti,ab OR
97. “floor effect”.ti,ab OR
98. “Item response model”.ti,ab OR
99. IRT.ti,ab OR
100. Rasch.ti,ab OR
101. “Differential item functioning”.ti,ab OR
102. DIF.ti,ab OR
103. “computer adaptive testing”.ti,ab OR
104. “item bank”.ti,ab OR
105. “cross-cultural equivalence”.ti,ab
106. COMBINE 1-105/OR
107. (Instrument’s name including acronyms, synonyms etc.)
108. COMBINE 106 & 107/AND

Individual searches were conducted for each of the instruments identified in Stage A in all three electronic databases.

(4) Stage B: Study selection

Articles returned from all of the searches were pooled and duplicates were removed. Study selection was undertaken following title selection, abstract selection and full article selection. Figure 8.2 displays a flowchart of Stage B processes including title selection, abstract selection, full article selection, data extraction and quality assessment. Title selection was conducted to exclude irrelevant and ineligible articles. Abstract selection checked if the study participants fitted the inclusion criteria and if the study evaluated the measurement properties of one or more of the identified PA instruments. In cases of any missing details or queries from the abstracts, the article was included for full article review. In Stage B piloting of study selection phases was not required as the process was replicated from that of Stage A. In Stage B the secondary reviewing team changed with the third reviewer (MH) being replaced by a new fifth reviewer (EH).

Figure 8.2 Flowchart of study selection process for Stage B.



(5) Stage B: Selection based on titles

Study selection on the title was conducted by the primary reviewer (RS); articles were only excluded if there was an explicit reason, for example, animal studies or studies in other populations, such as cardiac rehabilitation, this minimised reviewer error bias.

(6) Stage B: Abstract selection

All articles included for Stage B abstract selection were double-reviewed for eligibility by the primary reviewer and at least one of the secondary reviewing team (KD, GH & EH). If a reviewer noted missing detail or ambiguity, then the study was retained for full article screening. Each of the secondary reviewing team reviewed an equal share of the Stage B abstracts. Article screening forms were modified for Stage B (Appendix 1.7) from those used in stage A to check if articles assessed measurement properties of self-report PA instruments. After all abstracts had been reviewed the screening forms were compared for each abstract by the primary reviewer (RS). Disagreements between the two forms were resolved in a consensus meeting between the primary reviewer (RS) and the secondary reviewing team (KD, GH & EH) and a solution was found.

(7) Stage B: Full article selection

All of the articles selected for full screening were double reviewed by the primary reviewer and at least two of the secondary reviewing team (KD, GH & EH). Using the article selection form (appendix 1.7), all secondary reviewers (KD, GH & EH) submitted their completed screening to the primary reviewer (RS). Cases where there were disagreements in the screening forms were discussed in a consensus meeting between all of the reviewing team. Figure 8.2 provides a summary of all of

the selection processes for stage B. All included articles were then included for review of methodological quality and data extraction as described below.

(8) Stage B: Data extraction

Data extraction was conducted by: the primary reviewer (RS) and a combination of at least two of the secondary reviewers (KD, GH & EH) for each article to minimise reviewer and error bias. A data extraction form was designed specifically for this systematic review (Appendix 1.8), modified from the Quality Assessment of PA Questionnaires (QAPAQ) checklist (Terwee et al., 2010). The QAPAQ is a comprehensive checklist of all of the measurement properties and qualitative attributes of self-report PA instruments (Terwee et al., 2010). The data extraction form allowed reviewers to report information in a uniform manner.

The checklist was piloted using one eligible article (Svege et al., 2012). Part One of the QAPAQ extracts data on the instrument such as reporting the construct measure, recall period, purpose for the instrument. Appraisal of the instrument itself was conducted separately by the primary reviewer (RS) for each instrument identified. Changes to the wording of the QAPAQ were made by the primary reviewer (RS) following the pilot testing of the data extraction form for clarity. One item was changed from “Is the information being asked to evoke accurate answer?” to “Is the question asked in a manner to allow for an accurate answer?” Clear instructions as cues for the reviewers were also included in the data extraction form (Terwee et al., 2010).

(9) Stage B: Quality assessment

A quality assessment of the study methods highlights any study limitations or risk of bias. The quality assessment tool used was the COnsensus-based Standards

for the selection of health Measurement Instruments (COSMIN) checklist to evaluate articles assessing an instrument's measurement properties (Mokkink et al., 2010). The COSMIN 4-point rating scale was selected so that the grading of study quality could be used in the final synthesis of studies included in Stage B (Appendix 1.9). The COSMIN has been previously used in systematic reviews of PROMS, as part of the final grading of instruments for synthesis and findings (Schellingerhout et al., 2012; Saether et al., 2013; Weldam et al., 2013; Gilchrist et al., 2014). The COSMIN was modified for this systematic review: the item response theory, generalisability, and interpretability sections were removed. The latter two sections were removed as these data were collected in the QAPAQ, as part of data extraction. The item response theory section was removed, as all articles included in this systematic review used classical test theory to evaluate measurement properties. Such sections were then removed to reduce burden on the reviewing team. Box 8.3 summarises the modifications made to the COSMIN checklist.

Box 8.3 Summary of modifications made to COSMIN checklist for quality assessment stage.

Modifications made to COSMIN checklist
Removed sections on generalisability and interpretability of instruments
Removed section on item response theory
Items within the COSMIN that refer to item response theory methods removed in all of the sections
The term 'HR-PRO' changed to 'instrument' where used
Replaced term 'hypothesis testing' with 'construct validity'

Quality assessment was conducted alongside data extraction. Each article was quality assessed by the primary reviewer and one of the secondary reviewing team, to minimise human error and reviewer bias.

(10) Stage B: Qualitative evaluation

Qualitative evaluation of each instrument was conducted by the primary reviewer (RS) using a qualitative framework developed by the primary reviewer and adapted from part one of the QAPAQ (Terwee et al., 2010) and a descriptive framework based on the COSMIN definitions of measurement properties (Mokkink et al., 2010). The qualitative framework allowed reporting of the qualitative attributes of the instruments. Qualitative attributes of instruments used in this framework are summarised in table 8.5.

Table 8.5 Qualitative framework for Stage B data evaluation.

Qualitative Attribute	Definition
Construct	What construct of PA did the instrument intend to measure? e.g. Energy expenditure, daily activities, time spent walking
Setting	What settings are physical activities measured in? e.g. Work, leisure time, travel
Recall period	What is the recall period and length that the PA refers to? e.g Last week, usually week, month
Purpose	What is the purpose of the instrument? e.g. Clinical use, Research use
Target population	Who was the instrument originally developed for? e.g. Age, gender, health status
Justification	What was the rationale for the development of this instrument?
Format	Number of questions, answering format, scoring system.
Interpretability	Does the score give any clinical meaning? Are scores comparable to PA recommendations for health?

Ease of use	Is the time and effort to complete the instrument acceptable for the population it is intended? Are clear instructions included?

The primary reviewer assessed the qualitative attributes of all the self-report PA instruments. This was done by evaluating each instrument and was based on articles identified in Stage B that described the development of the instrument. The development and application of the instruments in any health research was described from published articles for each of the included instruments.

A descriptive framework was used to combine data on the measurement properties of the self-report PA instruments and the quality assessment of the methods used in the articles identified. This descriptive framework considered the evidence of measurement properties in 1) the musculoskeletal population: with joint pain or the OA population primarily and separately 2) the community dwelling population of adults aged 45 or over. The descriptive framework describes the evidence for each instrument and the strength of that evidence based on the quality assessment.

(11) Stage B: Grading of instrument measurement properties

A grading system was used for the descriptive framework along with the narrative described above, in order to rate the methodological quality of the measurement properties within articles identified in Stage B. The grading system gave a quantitative summary value of the evidence for each instrument's measurement properties, and the quality of that evidence. The grading system (Schellingerhout et al., 2012) has been used in other systematic reviews for assessing measurement properties of PROMS (Saether et al., 2013; Weldam et al., 2013;

Gilchrist et al., 2014). The scoring system included evaluating the strength of evidence for each measurement property in each instrument. The scoring in strength of evidence for overall quality of each measurement property is summarised in Table 8.6.

Table 8.6 Strength of evidence for overall quality of measurement property

Score in evidence strength	Rating	Criteria
Strong	+++ or ---	Consistent findings in multiple studies of good quality, or in one study of excellent quality
Moderate	++ or --	Consistent findings in multiple studies of fair quality, or in one study of good quality
Limited	+ or -	One study of fair quality
Conflicting	±	Conflicting findings among multiple findings
Unknown	?	Only studies of poor quality

The quality of the study was based on the COSMIN checklist for each of the measurement properties for each study, and was scored by the primary reviewer. Study quality was summarised for each measurement property of each study within the results of Stage B.

A positive or negative rating scored for the individual measurement properties was based on criteria from a list of measurement properties in each of the included articles (Terwee et al., 2007). The scoring system for the criteria in measurement properties is given in Table 8.7. The scores for each instrument in Stage B combined the strength of evidence with the positive or negative scores of the measurement properties, in order to provide a summary score in each measurement property.

Table 8.7 Criteria of measurement properties

Property	Definition	Rating	Quality Criteria
Criterion validity	The score of the instrument relates to that of the gold standard measurement	+	The gold standard used is double-labelled water, instrument's score correlates with gold standard ≥0.70
		?	Correlation not determined
		-	The gold standard used is double-labelled water, instrument's score correlates with gold standard correlation <0.70
		0	No information on criterion validity
Construct validity, compared with physical function or objective measure of PA	The degree to which scores from the instrument are related to measures of the same domain	+	Hypotheses are formulated, results are in accordance with hypotheses, and scores correlate ≥0.70 with measure of same domain
		?	No relationships determined
		-	Results are not in accordance with hypotheses or correlation <0.70
		0	No information on construct validity
Reliability	The extent to which scores between two or more populations can be separated to show true difference in score of a constructed	+	Intra-class correlation (ICC), correlation or weighted kappa ≥0.70
		?	No ICC, correlation, or weight kappa determined
		-	ICC, correlation or weight kappa <0.70
		0	No information on reliability

Internal consistency	The consistency of all the items within an instrument	+	Scale is unidimensional and Cronbach's alpha >0.69
		?	No Cronbach's alpha determined
		-	Scale is not unidimensional and Cronbach's alpha < 0.70
		0	No information on measurement error
Content validity	The extent that an instrument contains all facets of the construct measured	+	Target population considered all items relevant and questionnaire to be complete
		?	No target population involvement
		-	No target population considered all items relevant and questionnaire to be complete
		0	No information on content validity
Structural validity	The extent to which the dimensions of the instruments relate to the construct of interest	+	Factors explain at least 50% variance
		?	No factor analysis determined
		-	Factors explain < 50% variance
		0	No information on structural validity
Responsiveness	The ability of an instrument to detect change in the construct of interest	+	Change scores correlate with same constructs instruments >0.5
		?	Correlations in change scores not determined
		-	Change scores correlate with same constructs instruments <0.5
		0	No information on structural validity

The primary reviewer assessed the measurement property for each instrument in each study. This decided the direction of score for each instrument taken from Table 8.6. Criteria for adequate cross-cultural validity has not been previously recommended (Terwee et al., 2007; Schellingerhout et al., 2012). Thus, cross-cultural validity of instruments, where it was evaluated, is described within the narrative of the instrument.

8.5 Results

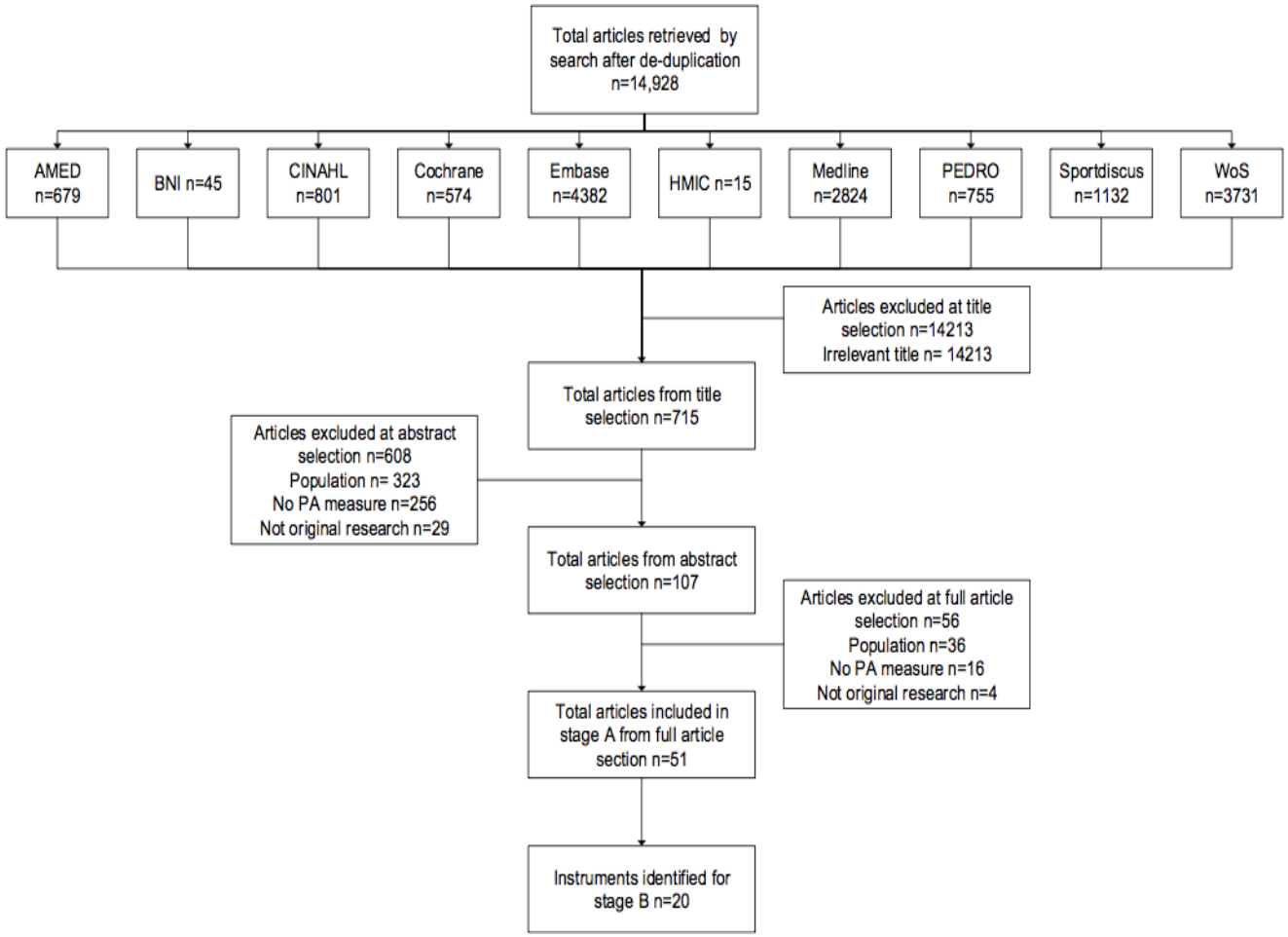
The results of the systematic review are reported in two stages. Stage A: identify self-report instruments using a systematic review of measures of PA previously used in adults aged 45 years and over with and without joint pain or OA. Stage B: describe the measurement properties of the self-report instruments identified by the systematic review measuring levels of PA in adults aged 45 years and over, with and without joint pain or OA.

8.5.1 Stage A: Identification of articles

From the search of electronic database, 15,104 articles were identified and after deduplication 14,928 articles were eligible for review (figure 8.3). By virtue of a comprehensive search strategy many of the articles were not relevant to the inclusion criteria or fitted the exclusion criteria. In Stage A the piloting of abstract and full article selection was conducted in 10 randomly selected articles, with 100% agreement achieved between all of the reviewing team (RS, KD, GH & MH) in both pilot scenarios with only minor changes requested by the secondary reviewing team, for example, refining the article selection form, rewording of items for clarity, and supplying definitions of measurement properties terms for quick

reference. An updated search of Stage A was conducted in the MEDLINE, EMBASE, WoS and Sportdiscus databases on the 18th of January 2016, using the same search strategy to identify additional articles. These databases were selected as they accounted for the majority of articles identified in the original search. A further 1,415 articles were identified in article searches. After screening titles, 177 (12.5%) articles were selected for abstract screening. There were 46 (26.0%) articles included for full-article screening. After this, an additional 36 articles were included for Stage A of the systematic review. Three new instruments were identified in the updated search; the Historical Leisure Activity Questionnaire (Jones et al., 2012), The Physical Activity Scale (Holm et al., 2014) and the Yale Physical Activity Survey (Chang et al., 2014).

Figure 8.3 Flow diagram of article selection for Stage A.



Key: Population = articles removed due to population did not fit inclusion criteria. No PA measure = No measurement of PA was made in the articles. Not original research = articles that were letter to editor or commentary.

Figure 8.3 displays the number of articles retrieved by each database search and the number of articles excluded at each step of selection. Articles were removed from the other databases if duplicates were found to those in MEDLINE, Embase and Web of Science databases. Following title review by the primary reviewer (RS) and review of abstract and full article by the reviewing team using the eligibility criteria, 51 articles were identified. The updated search identified an additional 36 articles, increasing the total number of articles in Stage A to 87.

8.5.2 Stage A: Identification of self-report PA instrument

Of 58 of the 87 studies, 29 analysed single items or contained no explicit detail on their content; therefore, these instruments were not analysed any further. Of the 23 instruments eligible for the next stage, 18 were multi-item self-reported PA questionnaires and 5 were single-item scale PA instruments.Table 8.8 summarises the instruments identified in Stage A and the studies of the instruments were identified.

Table 8.8 Identified instruments and eligible articles from Stage A.

Instrument	Eligible articles from Stage One
Multi-item Questionnaires	
Active Australia Survey (AAS)	Heesch et al. (2011)
Baecke Questionnaire	Ono et al. (2007) Terwee et al. (2011)
Daily Activity Questionnaire (DAQ)	Terwee et al. (2011) Wollmerstedt et al. (2010)
Historical Leisure Activity Questionnaire	Jones et al. (2012)
Human Activity Profile (HAP)	Davidson & De Morton, (2007) Terwee et al. (2011)
Incidental And Planned Activity Questionnaire For Older People (IPEQ)	Levinger et al. (2010)

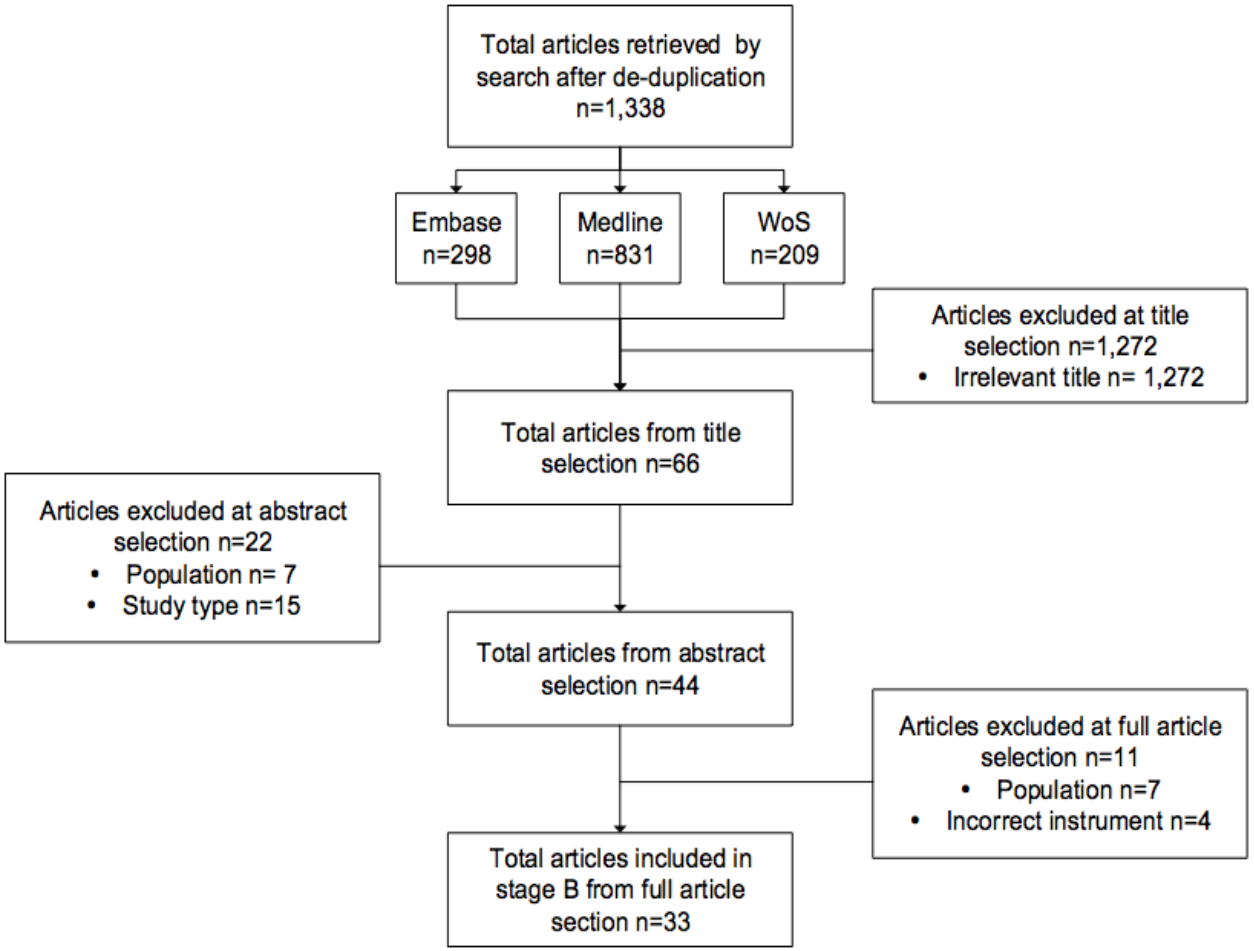
International Physical Activity Questionnaire (IPAQ)	Amer et al. (2014) Moreira-Silva et al. (2014) Naal et al. (2009, 2009a) Rosemann et al. (2007, 2007a, 2008, 2008a) Terwee et al. (2011) Svege et al. (2012) Tengman et al. (2014) Yazigi et al. (2013)
Longitudinal Ageing Study Amsterdam (LASA) Physical Activity Questionnaire	Verweij et al. (2009)
Minnesota Leisure Time Physical Activity Questionnaire (MLT-PAQ)	Ageberg et al. (2012) Barbour et al. (2014) Martin et al. (2013) Weller et al. (2006)
Modified Baecke Questionnaire	Thomas et al. (2003) Santos-Magalhaes & Hambly (2014)
The Physical Activity Scale	Holm et al. (2014)
Physical Activity Scale For Individuals With Physical Disablilities (PASIPD)	Groot et al. (2008)
Physical Activity Scales For The Elderly (PASE)	Batsis et al. (2014, 2015) Bennell et al. (2012, 2012a) Bossen et al. (2013, 2013a) Chmelo et al. (2014) Dunlop et al. (2011) Foster et al. (2014) Fransen et al. (2014) Hoogeboom et al. (2013) Lin et al. (2013) Mansourina et al (2012) Petrella et al. (2000) Hovis et al. (2011) Stehling et al. (2010) Svege et al. (2012, 2013) Bolszak et al. (2013)
Short Questionnaire To Assess Health Enhancing Physical Activity (SQUASH)	Kloek et al. (2014) Pister et al. (2010) Wagenmakers et al. (2008) Terwee et al. (2011)
The Adelaide Activities Profile (AAP)	Foley et al. (2003)
The Short Telephone Activity Rating Scale (STAR)	Holden et al. (2012, 2015)
Yale Physical Activity Survey	Chang et al. (2014)
Zutphen Physical Activity Questionnaire	Van Baar et al. (2001)
Single Item Scale	
Activity Rating Scale (ARS)	Santos-Magalhaes & Hambly (2014) Dawson et al. (2003) Naal et al. (2009, 2009a) Terwee et al. (2011)
Lower-Extremity Activity Scale (LEAS)	Terwee et al. (2011)

Tegner Scale	De Carvalho et al. (2014) Santos-Magalhaes & Hambly (2014) Naal et al. (2009) Terwee et al. (2011)
University Of California, Los Angeles Activity (UCLAA Scale)	Bauman et al. (2007) Fisher et al. (2011) Harding et al. (2014) Jensen et al. (2013) Lubbeke et al. (2014) Naal et al. (2009, 2009a) Terwee et al. 2011) Wollmerstedt et al. (2010)
Visual Activity Scale (VAS)	Terwee et al. (2011)

8.5.3 Stage B: Identification of articles

In Stage B searches of three electronic databases (MEDLINE, Embase and Web of Science) were conducted and 1,338 articles evaluated the measurement properties of the instruments identified in Stage A. Title, abstract and full review using the eligibility criteria for Stage B reduced this number to 33 articles. Figure 8.4 summarises the number of articles retrieved from the electronic databases and the number selected and excluded at each process. An updated search of databases was performed on the 18th of January 2016, using the same search strategy to identify additional articles for Stage B. A further 395 articles were identified in the article searches. After title and abstract screening of articles, 19 (4.8%) were retained in Stage B of the systematic review. Including the additional articles from the updated search, the total number of articles in Stage B was 52.

Figure 8.4 Flow diagram of article selection for Stage B.



Key: Population = articles removed due to population did not fit inclusion criteria, Study type = Article study did not assess measurement properties of self-report PA instrument, Incorrect instrument = A different instrument was used rather than one identified in Stage A.

The total number of articles included in Stage B was 52. Eight of the 52 (15.4%) articles, evaluated the measurement properties of the instruments in OA or joint pain population, while 44 (84.6%) of the 52 articles were conducted in community dwelling adult populations aged 45 and over. Of the 23 identified instruments, 13 had evaluated measurement properties in either an OA or joint pain population or a community dwelling adult population aged 45 years and over: IPAQ, PASE, Modified Beacke, Beacke, HAP, AAS, SQUASH, STAR, Tegner, UCLAA, ARS, IPEQ and the Zutphen Questionnaire. Table 8.9 summarises the instruments and the article(s) that evaluated the measurement properties.

Table 8.9 List of eligible articles in Stage B.

Instrument name	**Eligible articles from Stage Two**	
Multi-item Questionnaires	***Adult populations***	***OA/ joint pain populations***
Active Australia Survey (AAS)	Brown et al. (2008) Pettee et al. (2009) Heesch et al. (2011) Fjeldsoe et al. (2013) Winkler et al. (2013) Freene et al. (2014) Heesch et al. (2014)	NA
Baecke Questionnaire	Ono et al. (2007)	NA
Modified Baecke Questionnaire	Hertogh et al. (2008) Moore et al. (2008) Pols et al. (1995) Pols et al. (1996) Voorips et al. (1990)	NA
Daily Activity Questionnaire (DAQ)	NA	NA
Historical Leisure Activity Questionnaire	NA	NA
Human Activity Profile (HAP)	Bastone et al. (2014)	Bennell et al. (2004) Bilek et al. (2005)
Incidental and Planned Exercise Questionnaire for older people (IPEQ)	Delbaere et al. (2010) Merom et al. (2014)	NA

International Physical Activity Questionnaire (IPAQ)	Brown et al. (2004) Craig et al. (2003) Deng et al. (2008) Grimm et al. (2012) Heesch et al. (2011) Hurtig-Wennlof et al. (2010) Mader et al. (2006) Tomioka et al. (2011) Cerin et al. (2012) Dyrstad et al. (2014) Hansen et al. (2014) Kwak et al. (2012) Milanovic et al. (2013) Sebastiao et al. (2012) Tran et al. (2013) Van Holle et al. (2015)	Naal et al. (2009) Blikman et al. (2013)
Longitudinal Ageing Study Amsterdam (LASA) Physical Activity Questionnaire	NA	NA
Minnesota Leisure Time Physical Activity Questionnaire (MLT-PAQ)	NA	NA
Physical Activity Scale for Individuals with Physical Disablilities (PASIPD)	NA	NA
Physical Activity Scales for the Elderly (PASE)	Colbert et al. (2011) Dinger et al. (2004) Hagiwara et al. (2008) Harada et al. (2001) Moore et al. (2008), Schuit et al. (1997) Washburn & Ficker (1999) Washburn et al. 1993) Ngai et al. (2012) Vaughn & Miller (2013)	Martin et al. (1999) Svege et al. 2012) Bolszak et al. (2012) Casartelli et al. (2015)
The Physical Activity Scale (PAS)	NA	NA
Short Questionnaire to Assess Health-Enhancing Physical Activity (SQUASH)	De Hollander et al. (2012).	NA
The Adelaide Activities Profile (AAP)	NA	NA
The Short Telephone Activity Rating scale (STAR)	Matthews et al. (2005)	NA
Yale Physical Activity Survey	NA	NA
Zutphen PA Questionnaire	Harris et al. (2009)	NA
Single Item Scale	***Adult populations***	***OA/ joint pain populations***
Activity Rating Scale (ARS)	NA	Naal et al. (2009)

Lower-Extremity Activity Scale (LEAS)	NA	NA
Tegner scale	NA	Naal et al. (2009)
University of California, Los Angeles Activity (UCLAA) scale	NA	Naal et al. (2009)
Visual Activity Scale (VAS)	NA	NA

Key: NA = No articles retrieved evaluating measurement properties of instruments in the population.

8.5.4 Evaluation of individual instruments

This section provides details of each individual instrument where measurement properties have been evaluated. Qualitative attributes for each of the 13 instruments are given as well as measurement properties in both OA or joint pain populations and community dwelling adult populations aged 45 and over.

Active Australia Survey (AAS)

Table 8.10 provides a summary of the qualitative attributes of the AAS (Armstong et al., 2000). The AAS was originally developed for adults aged 18-65 years old but has been used in populations that have ranged above 65 years old (Armstong et al., 2000). The AAS contains items on leisure time walking, walking for working purposes, vigorous exercises or activities and moderate exercises or activities; the AAS has a 7 day recall for all of these items. The AAS can categorise individuals according to PA guidelines for time spent in the last week moderately active or vigorously active for health benefits.

Table 8.10 Qualitative attributes of the AAS.

Qualitative Attribute	Definition
Construct	Leisure time PA
Setting	leisure time walking, walking for working purposes, vigorous exercises or activities and moderate exercises or activities
Recall period	7 Days
Purpose	To assess knowledge of health benefits of PA in adult populations
Target population	Developed for adults aged 18-65, can be used internationally
Justification	Offers data on PA that can be implemented into self-report survey or interviewing
Format	Nine items, entering time spend during activities or frequency of activities
Interpretability	Total time spent during a week PA and time spent sedentary
Ease of use	Short time taken to complete

(1) Reliability

Three studies were identified evaluating the measurement properties of the AAS, all of these studies were in community adult populations aged 45 and over (Brown et al., 2008; Pettee et al., 2009; Heesch et al., 2011). No articles were retrieved that assessed the AAS in adult populations with joint pain or OA. Two of these studies evaluated reliability: Pettee et al. (2009) found an insufficient test-retest reliability in the AAS (Spearmans rank, r=0.32) in an adequate sample (n=66). Brown et al. (2004) found a higher but still insufficient reliability for frequency of activity per week (Spearmans rank, r=0.64) and for minutes of activity per week (Spearmans rank, r=0.58) in a larger sample (n=154).

(2) Construct Validity

In one of the studies, construct validity was assessed comparing the AAS score to that of total activity count on an Actigraph activity monitor (Pettee et al., 2009). The findings demonstrated moderate correlations between the AAS scores and Actigraph activity count (r= 0.39). This study together with another study (Heesch et al., 2011) correlated the AAS score with pedometers and moderate correlations were found (r= 0.42), which was similar in Pettee et al. (2009) for correlations with pedometers (r=0.49). Table 8.11 summarises the measurement properties of the AAS, methodological quality was reported in brackets in each measurement property for each study.

(3) Updated search

Four studies were identified in the updated search that assessed the measurement properties of the AAS, one of which evaluated a modified version of the AAS. The modified version of the AAS was identical to the original but with four additional items added in each domain, asking individuals to indicate how many days per week they engaged in each activity.

Fjeldose et al. (2012) evaluated reliability and construct validity of the modified AAS in 63 adults, either university staff or students between the ages 30-70 years. Mean age of the participants was 49.5 years and 36.5% were males. The test retest reliability evaluation was conducted during a 3-5 day interval, despite the AAS having a 7 day recall period. Reliability was reported in domains rather than total scores. The moderate and vigorous activity scores were highly correlated using Spearman's rank correlations (r=0.80). Time spent walking was also highly

correlated using Spearman's rank correlation (r=0.76). The modified AAS's additional items demonstrated kappa agreement. For days of moderate and vigorous activities the kappa was 0.63. For days of walking activities the kappa was 0.64. Construct validity was assessed comparing the modified AAS score with an Actigraph GT1M accelerometer. Correlation between the AAS and Actigraph was strong for moderate and vigorous activities (r=0.61). Agreement between days of moderate and vigorous activities and Actigraph was (k=0.45).

Freene et al. (2013) evaluated construct validity of the original AAS in adults that were enrolled onto a community based exercise programme (n=37) or home based exercise programme (n=37) (. Mean ages of the community based exercise group and the home based exercise group were 59.9 and 56.7 years of age respectively. Construct validity was evaluated by correlating the AAS scores with an exercise diary and the ActiGraph 5s. In the community based exercise group, the AAS score demonstrated moderate correlation with the Actigraph 5s (r=0.49) and a strong correlation with the exercise diary (r=0.64). In the home exercise group the correlations were similar, with the Actigraph 5s demonstrating a correlation of r=0.56 and the diary demonstrating a correlation of r=0.56.

Heesch et al. (2014) evaluated the understanding and interpretation of the AAS in 55 adults aged 65 years and over. The study aimed to identify how individuals understand and interpret the items of the AAS using a cognitive interview approach. A large number of wide ranging issues were identified with the AAS. The results of the study identified issues with the scope of the activities included within the different domains of the AAS. There were also issues for participants regarding the interpretation of 'time' and 'number of times' in the items of the AAS,

where at least 10 minutes was considered a cumulative value rather than per bout of activity. There were also issues where responders duplicated the same activity across domains of the AAS. Heesch et al. (2013) noted that these issues with the ASS where also common in other PA self-report instruments.

Winkler et al. (2013) compared measurement error in the AAS in an exercise intervention programme for weight loss and a non-exercise control group in adults recruited from primary care (mean age 58.3±8.6 years). The study reported that at 6 month follow up the intervention group reported significantly lower measurement error compared to control group. The study also reported that in the control group there was a significant difference in PA levels between the AAS and Actigraph GT1M.

Table 8.11 Summary of the AAS measurement properties

Article	Population	Sample size	Reliability	Criterion validity	Construct validity using objective measure	Construct validity using physical function	Content Validity
Brown et al. (2008)	General female population, mean age 55 years	N=154	Spearmans rank r=0.58 for frequency/ week, 0.64 for minute/ week (fair quality)	0	accelerometer, Spearmans rank r=0.48 for for frequency/ week, 0.52 for minutes per week (good quality)	0	0
Heesch et al. (2011)	Community dwelling adults, mean age 72 years	N=53	0	0	Correlation to pedometers, 0.42 (good quality)	Correlation to physical component of SF-36, 0.39 (fair quality)	0
Pettee et al. (2009)	General female population, mean age 52 years	N=66	Spearmans Rank, r=0.32 (fair quality)	0	Correlation to Actigraph accelerometer, 0.39, Correlation to pedometer, 0.49, (good quality)	0	0
Freene et al. (2013)	General adult population, mean age 56.7-59.9	N=37 N=37	0	0	Correlation to Acitgraph 5s (0.49, 0.56) (good quality)	Correlation to exercise diary (0.64, 0.56) (fair quality)	0

Fjeldose et al. (2012)	University staff or students, mean age 49.5 years	N=63	Spearmans Rank, r=0.76, r=0.8 kappa 0.63, 0.64 (good quality)	0	Correlation to Actigraph accelerometer, r=0.61, agreement with number of days items 0.45 (good quality)	0	0
Heesch et al. (2014)	General adult population, mean aged 65 years and over	N=55	0	0	0	0	Wide range of limitations in items (excellent quality)
Winkler et al. (2013)	Primary care adults, mean age 58.3 years	N=302	Sig. difference in intervention and control for measurement error (poor quality)		NS difference between Actigraph and intervention group, sig. dif between Actigraph and control (good quality)		

Key: Excellent quality, good quality, fair quality refers to the methodological quality of the study measurement property assessed using the COSMIN checklist (Mokkink et al., 2010), 0 = no evaluation of the measurement property within the study.

Baecke questionnaire

The Baecke questionnaire is a self-report PA instrument developed for use in epidemiological studies to assess levels of PA in young adults (Baecke et al., 1982). The instrument contains three domains: work related, leisure time and sport activities; containing 16 closed answered items. Table 8.12 provides a summary of qualitative attributes of the Baecke PA questionnaire (Baecke et al., 1982).

Table 8.12 Qualitative attributes of the Baecke PA questionnaire

Qualitative Attribute	Definition
Construct	Habitual PA across three domains; work related activity, leisure time activity and sport
Setting	Activities in: occupation, movement, sport, leisure time activities that excluded sport, and sleeping habits.
Recall period	Usual week
Purpose	To assess habitual physical activities for epidemiological studies
Target population	Young adults
Justification	At the time of development no appropriate instrument was available for use in epidemiological studies
Format	Self-report questionnaire with 16 items closed answered questions
Interpretability	Scores are given in three indices; work, sport, leisure time. These scores are not interpretable outside of the Baecke
Ease of use	Small number of multiple choice questions

(1) Reliability and construct validity

The systematic search only identified one article assessing the measurement properties of the Baecke questionnaire. The study assessed the Baecke

questionnaire in terms of its reliability and construct validity against pedometer activity monitors in a sample of males and females with hip OA in one or more joint (n=51) (Ono et al. 2007). The Baecke questionnaire was shown to be reliable with a high ICC (0.87) and a moderate correlation to pedometer counts (r=0.49). Table 8.13 summarises the measurement properties of the Baecke questionnaire, methodological quality was reported in brackets in each measurement property for each study.

Table 8.13 Summary of the Baecke questionnaire measurement properties

Article	Population	Sample size	Reliability	Criterion validity	Construct validity using objective measure	Construct validity using physical function
Ono et al. (2007)	Hospital outpatients, females with hip OA, mean age 53	N= 51	ICC= 0.87 (good quality)	0	Correlation to accelerometer, 0.49 (good quality)	0

Key: Excellent quality, good quality, fair quality refers to methodological quality of study in each measurement property, quality was assessed using the COSMIN checklist (Mokkink et al., 2010), 0= no evaluation of measurement property within the study

Modified Baecke

The modified Baecke is a short interview-based PA measure, based on the Baecke questionnaire (Voorrips et al., 1990). The Baecke was modified for use by the elderly and adapted for interviewer administration; it takes approximately 30 minutes to complete (Voorrips et al., 1990). The recall period of the modified Baecke is one year, responders are asked to report physical activities in the household and in sports or other activities and report time spent in hours over a week; the modified Baecke is scored by weighting activities depending on the energetic cost of those activities. A summary of the qualitative attributes of the modified Baecke is provided in table 8.14.

Table 8.14 Qualitative attributes of the Modified Baecke

Qualitative Attribute	Definition
Construct	Physical activities in household and sporting activities
Setting	Household activities and leisure sporting activities
Recall period	One year
Purpose	Modified to better suite elderly population from the original Baecke
Target population	Elderly adults, aged 65 years and over
Justification	Original Baecke not appropriate for elderly populations.
Format	Interviewer administered not self-report like the original Baecke
Interpretability	Time spent PA in hours during one week. Scores can be compared to recommendations on PA levels for health benefits
Ease of use	Interviewer required, takes approximately 30 minutes to complete.

(1) Reliability

No study evaluating the modified Baecke for joint pain or OA populations was found. Reliability was evaluated in three studies; in one study reliability was assessed over a 20 day interval and although sample size was small (n=29), the modified Baecke was found to be reliable (ICC=0.89) (Voorrips et al., 1990). In another study, five month and 11 month reliability was tested in a sample of adult participants (n=35) (Pol et al., 1996); the modified Baecke was found to be reliable at both time points (Pol et al., 1996). The same research group also found that the modified Baecke was reliable in a larger sample of males and females (n=126) at five months (Spearmans rank= in males 0.85 and females 0.83) and at 11 months (Spearmans rank= in males 0.83 and females 0.77) (Pol et al., 1995).

(2) Criterion validity

One study evaluated the criterion validity of the modified Baecke with the DLW method in a small sample (n=21) (Herogh et al., 2008). A moderate Spearman's rank correlation (r=0.54) was found between modified Baecke scores and energy expenditure measured with the DLW method.

(3) Construct Validity

Construct validity was evaluated for the modified Baecke comparing scores with self-report exercise diaries (Voorrips et al., 1990; Pols et al., 1995 & 1996). One study found a strong association (r=0.78, Voorrips et al., 1990), while others found moderate associations (r=0.44-0.56) (Pols et al., 1995 & 1996). Pols et al. (1996) found a non-significant correlation between modified Baecke scores and heart rate monitoring in a small sample (n=35). Moore et al. (2008) compared the modified

Baecke to the PASE in 54 elderly adults and non-significant correlations were found. Table 8.15 summarises the measurement properties of the modified Baecke questionnaire, methodological quality was reported in brackets in each measurement property for each study.

Table 8.15 Summary of the modified Baecke questionnaire measurement properties

Article	Population	Sample size	Reliability	Criterion validity	Construct validity using objective measure	Construct validity using physical function
Herogh et al. (2008)	Elderly population, mean age 69 years	N=21	0	Correlation with DLW, r=0.54 (poor quality)	0	0
Moore et al. (2008)	General population over 50, mean age 70	N=54	0	0	0	Non- significant correlation with PASE (good quality)
Pols et al. (1995)	Adults population, mean age 41-48 years	N= 126	Spearman rank, r=0.65-0.89 (fair quality)	0	0	0
Pols et al. (1996)	Community dwelling females, mean age 61 years	N= 35	Pearson's correlations, 5 months= 0.82, 11 months= 0.73 (poor quality)	0	Correlation to heart rate monitoring, not significant (poor quality)	0
Voorips et al. (1990)	General population, mean age 69-73 years	N= 29	Spearman rank, r=0.86 (poor quality)	0	Correlation to pedometer, r=0.72 (poor quality)	0

Key: Excellent quality, good quality, fair quality refers to the methodological quality of the study measurement property assessed using the COSMIN checklist (Mokkink et al., 2010), 0 = no evaluation of the measurement property within the study.

Human activity profile (HAP)

The HAP differs from other PA questionnaires; originally developed as an indicator of quality of life in those with chronic obstructive pulmonary disease (COPD) after pulmonary rehabilitation (Bennell et al., 2004). The HAP consists of 94 items which are a list of daily activities organised in ascending order due the amount of energetic output required to perform the activity. Responders indicate on the list if they can perform the activity unassisted, if they were able to perform the activity before but now cannot, or if they never did the activity. From this responders can be scored in physical function ability or maximal activity score (MAS) and average level of PA or adjusted average activity (AAS) can be derived; scores are classified into certain levels of PA (Bennell et al., 2004). The HAP only measures intensity of PA and does not measure duration of activity although this can be estimated based on the intensity that individual's report they are able to achieve. Table 8.16 provides a summary of qualitative attributes of the HAP (Bennell et al., 2004, Davidson & De Morton 2007).

Table 8.16 Qualitative attributes of the HAP.

Qualitative Attribute	Definition
Construct	Energy expenditure or physical fitness
Setting	Daily activities
Recall period	Same day
Purpose	Originally developed as indicator of quality of life in pulmonary rehabilitation
Target population	Clinical and healthy populations
Justification	Previously developed instruments were developed for a too specific of a population and had floor and ceiling effects
Format	List of 94 items, each one a daily activity,
Interpretability	Scores give average levels of activity and maximal achievable activity, although does not give duration, frequency or intensity of activities for scores to be interpretable
Ease of use	Closed answer questions, time taken to complete: 1-2 minutes

(1) Reliability

Two studies evaluated the measurement properties of the HAP, one study included participants with knee osteoarthritis (n=226) (Bennell et al., 2004) and the other study included females with either OA or rheumatoid arthritis (RA) (OA n =16, RA n=12) (Bilek et al., 2005). Both studies assessed reliability with Bennell et al. (2004) finding high reliability in both scoring methods (MAS ICC=0.96, AAS ICC=0.95) and a small standard measurement error (SEM=3). The second study found a smaller value for reliability (MAS ICC=0.60, ASS ICC=0.83), which may be explained by the smaller sample size (Bilek et al., 2005).

(2) Construct validity

Both studies described above also found weak to strong correlations with physical function (Bennell et al., 2004; Bilek et al., 2005), Bennell et al. (2004) found weak correlations with the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) (r=-0.23-0.39) and moderate correlations in observed physical function tests (r=0.34-0.63). Bilek et al. (2005) found stronger correlations between HAP scores and self-report physical function measured by the SF-36, (r=0.78-0.80). Table 8.17 summarises the measurement properties of the HAP, methodological quality was reported in brackets in each measurement property for each study.

(3) Updated search

One study was identified in the updated search that evaluated measurement properties in the HAP (Bastone et al., 2014). The study aimed to evaluate construct validity of the HAP in community dwelling female adults aged 60 years and over (n=133). Construct validity was assessed by comparing the HAP to the Actigraph GT3X activity monitor. A significant and moderate strength correlation between the Actigraph and the AAS was demonstrated (0.52) and MAS (0.55) score of the HAP was reported.

Table 8.17 Summary of the HAP measurement properties

Article	Population	Sample size	Reliability	Criterion validity	Construct validity using objective measure	Construct validity using physical function
Bennell et al. (2004)	Community adults with knee OA, mean age 69 years	N=226	MAS ICC= 0.96, AAS= 0.95, SEM=3 (excellent quality)	0	0	Correlation to WOMAC physical function, MAS=-0.23, AAS= -0.39 (good quality)
Bilek et al. (2005)	Radiographic OA, mean age 60 years	N= 16	MAS ICC= 0.6 AAS ICC= 0.83 (poor quality)	0	0	Correlation with SF-36, MAS= 0.78, AAS= 0.8 (poor quality)
Bastone et al. (2014)	Community dwelling female adults aged 60 and over, mean age 71.8	N=133	MAS ICC= 0.79, AAS= 0.94 (good quality)	0	Correlation with Actigraph GT3X, AAS=0.52, MAS=0.55 (good quality)	0

Key: Excellent quality, good quality, fair quality refers to the methodological quality of the study measurement property assessed using the COSMIN checklist (Mokkink et al., 2010), 0 = no evaluation of the measurement property within the study.

Incidental and planned exercise questionnaire (IPEQ)

Two versions of the IPEQ are available, a past week version and an average week over a three month period version. The IPEQ contains 10 items on planned and structured exercise and contains items on exercise as part of daily life, these are scored to give an overall score for the duration of a week (Delbaere et al., 2009). Table 8.18 provides a summary of qualitative attributes of the IPEQ (Delbaere et al., 2009).

Table 8.18 Qualitative attributes of the IPEQ questionnaire.

Qualitative Attribute	Definition
Construct	Incidental and planned physical activities
Setting	Gym or home, activities in daily life
Recall period	7 days or 3 months
Purpose	Used in longitudinal epidemiology studies to assess levels of PA
Target population	Frailer populations
Justification	Other instruments for adults aged 45 years and over have too many items for survey use
Format	Two parts; planned or structured exercises and activities in daily living
Interpretability	Scores are interpretable to time spent physically active
Ease of use	10 short items, self-complete instrument, quick to complete

(1) Reliability

Only one study evaluated the measurement properties of both versions of the IPEQ in a sample of elderly adults aged 70 years and over in both the past week

version (n= 230) and average week version (n=230) (Delbaere et al., 2010). Reliability was the only measurement property evaluated in this study and the recall period was seven days. Given that one version of the IPEQ had recall for an average week over the past three months; this may have biased the reliability analysis with overlapping recall periods. Findings for reliability suggested that both versions were sufficiently reliable (past week ICC=0.84, average week ICC= 0.80). Table 8.19 summarises the measurement properties of the IPEQ, methodological quality was reported in brackets in each measurement property for each study.

(2) Updated search

One new study was identified in the updated search that evaluated construct validity and responsiveness of the IPEQ in adults aged 65 and over (Merom et al., 2014). Construct validity was evaluated comparing the IPEQ scores to an ActiGraph GT1M and responsiveness was evaluated by comparing the responsiveness index (mean change of intervention/SD of control) of the IPEQ to the ActiGraph. Correlations between the ActiGraph and IPEQ were low (r= 0.17), suggesting poor construct validity. The responsiveness index of the IPEQ was much lower (0.31) compared to the ActiGraph (0.65) showing poor measurement of change in the IPEQ.

Table 8.19 Summary of the IPEQ measurement properties

Article	Population	Sample size	Reliability	Criterion validity	Construct validity using objective measure	Construct validity using physical function	Responsivness
Delbaere et al. (2010)	Community adult sample, mean age 77	N= 230	ICC= 0.84 average week, ICC=0.80 past week (good quality)	0	0	0	0
Merom et al. (2014	Adults aged 65 and over	N=315	0	0	Low correlation to ActiGraph GT1M, 0.17 (excellent quality)		IPEQ responsiveness index 0.31, ActiGraph 0.65 (excellent quality)

Key: Excellent quality, good quality, fair quality refers to the methodological quality of the study measurement property assessed using the COSMIN checklist (Mokkink et al., 2010), 0 = no evaluation of the measurement property within the study.

International Physical Activity Questionnaire (IPAQ)

Table 8.20 introduces the qualitative attributes of the IPAQ (Booth et al., 2003; Craig et al., 2003). The IPAQ is a self-report PA instrument developed to be used in any population and is the first self-report physical questionnaire that can be used internationally. As a generic self-report outcome measure, different clinical and non-clinical populations can be compared. The IPAQ was developed by an international expert consensus meeting producing a self-report PA questionnaire suitable for assessing levels of PA in populations across different countries (Booth et al., 2003). Pilot testing was also part of developing the IPAQ, this was not explicitly described (Craig et al., 2003). There are two different versions of the IPAQ available: a short version (IPAQ-SF), and a long version (IPAQ-LF), both can be either self-administered or telephone administered (Craig et al., 2003).

The IPAQ has many translations available for different cultures (www.ipaq.ki.se). The short version of the IPAQ contains four generic items: two on vigorous and moderate PA, one on walking and one on time spent sitting within the last 7 days (Booth et al., 2003). The long version of the IPAQ is split into five separate domains of PA including: job related PA, PA for transport, housework activities, leisure time sports and time spent sitting (Craig et al., 2003). The score for the IPAQ is given as a continuous score in the metabolic equivalent (MET) which is a measure of PA intensity per minute, per week ($METS^{-1min\text{-}1week}$), this allows scores to be directly interpreted into the frequency, duration and intensity of PA over a week and be compared to recommendations for levels of PA.

Table 8.20 Qualitative attributes of IPAQ.

Qualitative Attribute	Definition
Construct	Energy expenditure in a week. There is a long version and short version
Setting	Long version includes; transport, housework activities, leisure time sports and time spent sitting. Short version does not separate into different settings
Recall period	Two versions; last week and usual week
Purpose	Research to compare populations in levels of PA
Target population	Adults, 18-65 years old. Different languages available
Justification	A generic outcome measure of PA to be used in any adult population internationally
Format	Short version: 4 items, Long version: 5 domains, 27 items. Closed questions, some with continuous scale answer
Interpretability	Scores given in energy expenditure per week, scores can be compared to recommendations on PA levels for health benefits
Ease of use	Short version requires minimal time and effort. Long version takes longer and requires recall in different aspects of PA

(1) Reliability

One study was found which evaluated measurement properties for the IPAQ-SF in OA and joint pain populations (Naal et al., 2009). This study tested the reliability and measurement error of the IPAQ-SF in a pre-total hip or knee surgery sample aged between 34-88 years old (n=43 hip OA and n=36 knee OA) and was administered twice in a one week interval. The IPAQ-SF was found to be reliable in both knee OA (ICC= 0.87, 95% CI, 0.74–0.94) and hip OA (ICC= 0.76, 95% CI, 0.57–0.87), with values above the recommended 0.70 (Terwee et al., 2007).

Although sample sizes were smaller than the recommended 50, both samples combined showed positive reliability in a sample greater than 50.

In community dwelling adults aged 45 years and over, eight studies were found evaluating measurement properties of the IPAQ. Five of these studies evaluated the reliability of the IPAQ in community dwelling adult populations, 4 investigated the IPAQ-SF (Brown et al., 2004; Deng et al., 2008; Mader et al., 2006; Tomioka et al., 2011). Craig et al., (2003) investigated the IPAQ-LF and IPAQ-SF. Two of these studies displayed insufficient reliability for the IPAQ (Brown et al., 2004; Tomioka et al., 2011). One study also found insufficient correlation using a Spearman's rank correlation test (r=0.54) (Mader et al., 2006). Two studies, including an international multicentre study, found positive scores for reliability with sufficient ICCs or Spearmans rank correlations above 0.70 in sample sizes above n=50 (Deng et al., 2008, Craig et al., 2003).

(2) Construct Validity

Six studies evaluated construct validity in the IPAQ-SF in community dwelling adult populations using comparisons with objective measures of PA (Craig et al., 2003; Mader et al., 2006; Deng et al., 2008; Hurtig-Wennlof et al., 2010; Tomioka et al., 2011; Grimm et al., 2012) and one in the IPAQ-LF (Craig et al., 2003). Five studies found a weak to moderate correlation of the IPAQ-SF score with the accelerometer measure; Tomioka et al. (2011) did not identify which accelerometer was used but found a moderate correlation with IPAQ scores (r=0.5-0.65). Mader et al. (2006) found a moderate correlation using the Actigraph activity monitor (r=0.54), although the sample size in this study was small (n=35). Hutig-Wennlof et al. (2010) found a weak Kappa agreement with IPAQ-SF and the

Actigraph activity monitor categorising individuals into different levels of PA (r=0.45). Grimm et al. (2012) found a weak Spearman's correlation (r=0.24) with energy expenditure using the Actigraph activity monitor. Craig et al. (2003) found moderate correlations in the objective measures of PA in their international multicentre study with both IPAQ-SF and IPAQ-LF (r=0.23-0.39). One study also found moderate correlations in IPAQ-SF scores with pedometers (Deng et al., 2008) (r=0.33) and Brown et al. (2004) found moderate correlations with another self-report PA instrument, the AAS (r=0.68).

(3) Content validity

One study (Heesch et al., 2011) evaluated whether the IPAQ-SF was appropriate for adult populations aged 65 years and over. Using qualitative interviewing Heesch et al. (2011) investigated the IPAQ-SF's items in elderly populations. They found that the definitions used for different intensities of exercise were confusing to adults and recall of activities lasting at least 10 minutes was difficult for this population (Heesch et al., 2011). Table 8.21 summarises the measurement properties of the IPAQ-SF and IPAQ-LF, methodological quality was reported in brackets in each measurement property for each study.

(4) Updated search

In the updated search; 8 articles were identified to have evaluated the IPAQ-SF (n = 3) and IPAQ-LF (n=6). One study evaluated construct validity and reliability of the IPAQ-SF and IPAQ-LF in an adult population that either had a total knee replacement or total hip replacement due to OA (Blikman et al., 2013). In this study there were 44 participants (47% with total knee replacement, and 43% with

total hip replacement) with a mean age of 72±9. Construct validity was evaluated by comparing both versions of the IPAQ to an ActiGraph GT1M activity monitor. The IPAQ-LF had a stronger association to the ActiGraph GT1M activity monitor (r=0.43) compared to the IPAQ-SF (r=0.29). The IPAQ-LF was also found to be more reliable in test-retest (ICC=0.65) compared to the IPAQ-SF (ICC=0.51). Both reported high standard error of measurement (SEM) and smallest detectable change (SDC) (IPAQ-LF, SEM=2668, SDC=1115; IPAQ-SF, SEM=2487, SDC=1039).

Cerin et al. (2012) evaluated an interview administered, rather than self-report Chinese version of the IPAQ-LF in a sample of 94 adults based in Hong Kong aged 65 years and over (42% males), mean age was not reported. Construct validity was tested by correlating IPAQ-LF with an ActiGraph GT1M activity monitor, reliability was evaluated using test-retest methods. Total IPAQ-LF scores moderately correlated with activity monitor total PA scores (r=0.39), high reliability was reported in the IPAQ-SF (ICC=0.86).

A study in Norway evaluated the construct validity of the IPAQ-SF in a large sample of community dwelling adults (n=1751, mean age 48.2) (Dyrstad et al., 2014). Construct validity was evaluated by comparing IPAQ-SF scores to an ActiGraph GT1M activity monitor and a moderate correlation (r=0.33) was found.

Hansen et al. (2014) evaluated the Danish version of the IPAQ-LF for reliability and construct validity compared to an activity monitor that collected both activity monitor data and heart rate data to predict energy expenditure (Actiheart sensor). Study participants were 121 healthy adults (42% males) with a mean age of

49.0±13.2 years. Total scores of the IPAQ-LF had significant but weak associations with total energy expenditure in the Actiheart sensor for females (n=64, r=0.25) and moderate associations for males (n=57, r=0.46). Overestimations in the IPAQ-LF were reported in the moderate and vigorous activities domain of the IPAQ-LF compared to the Actiheart sensor in both males and females. Reliability of the IPAQ-LF was reported as low (ICC=0.58), as the score was below the recommended ICC of 0.7 guideline cut-off.

Milanovic et al. (2014) evaluated a Serbian translated version of the IPAQ-LF for reliability in adults aged 65 and over. The translation was performed from the English version of the IPAQ-LF by two independent translators, who were familiar with PA questionnaires. Changes were reported to have been made to the final Serbian version, although these changes were not described in detail. The Serbian version was then translated back to English and checked. A 14 day test-rest format was conducted in a sample of 660 adults with a mean age of 67.65±5.76 years. The ICC of the Serbian version of the IPAQ-LF was 0.74, demonstrating good reliability.

Sebastião et al. (2012) explored issues associated with measuring PA using the IPAQ-LF in a Brazilian adult population. Study participants were adults (657 males, 915 females) with a mean age of 44.8±17 years for males and 46.7±17 years for females. Scores of the IPAQ-LF were broken down into their domains and explored for unusual distributions of PA. The results demonstrated issues with

overestimations of PA in work-related and household-related activity compared to expected levels of PA in Brazilian adults.

Tran et al. (2013) evaluated a Vietnamese translated version of the IPAQ-SF. The translation of the IPAQ-SF was translated and back-translated by an independent translator with original meaning of the items kept but changed to fit culturally. Face and content validity of the translated version were reported as tested in a pilot; however this was not explicitly explained within the study. The study evaluated validity of the translated IPAQ-SF compared to a Yamax SW-200 pedometer and an activity logbook. Reliability was assessed on a 3-day test-retest approach, although the IPAQ-SF is a 7 day recall questionnaire, this meant that there was an overlap in the time frame of the reliability evaluation. The translated version of the IPAQ-SF was evaluated in 150 (50% males) adults aged 60 and over and a mean age of 66.8±5.1 years. There was a moderate correlation between the logbook and IPAQ-SF (r=0.46) and a weak correlation was found with the pedometer (r=0.20). Reliability was high with an ICC of 0.91, although the test-retest period was only 3 days.

Van Holle et al. (2015) evaluated a modified-Belgian version of an interview administered IPAQ-LF for adults aged 65 and over. The IPAQ-LF was modified by combining moderate and vigorous PA into one domain, a measurement of gait speed was added to the walking domain and a domain on cycling was added to represent a European lifestyle. The modified IPAQ-LF was evaluated in adults aged 65 and over for construct validity; using comparisons to an ActiGraph GT3X activity monitor (n=434). In a subsample; test-retest reliability was evaluated (n=29), mean age of the sample was 74.2±6.2 year (53.7% females). In terms of

validity the IPAQ-LF correlated with a Copeland cut-off point for PA of 0.36 and a Freedson cut-off point for PA of 0.40. Test-retest reliability was conducted with a mean time-gap of 9.6 days. The ICC in the reliability subsample was ICC=0.63.

Table 8.21 Summary of the IPAQ measurement properties

Article	Population	Sample size	Reliability	Criterion validity	Construct validity using objective measure	Construct validity using physical function	Other measurement properties
Naal et al. (2009) (IPAQ-SF)	Total knee + hip replacement patients, mean age 63-66 years	N= 105	ICC=0.87 knee replacement, ICC=0.76 (excellent quality)	0	0	0	0
Brown et al., (2008) (IPAQ-SF)	General population, age range 18-75 years	N= 104	ICC=0.68 (excellent quality)	0	0	0	0
Craig et al. (2003) (IPAQ-SF & IPAQ-LF)	International study, mean age range 46-56 years	N=725	Spearman rank, r=0.46-0.96 (good quality)	0	Correlation to accelerometer, r=0.30-0.33 (good quality)	0	0
Deng et al. (2008) (IPAQ-SF)	Elderly general population, mean age 63-67 years	N= 224	ICC= 0.84 (excellent quality)	0	Correlation to pedometer, r=0.33 (fair quality)	0	0
Grimm et al. (2012) (IPAQ-SF)	General population, mean age 64 years	N= 127	0	0	Correlation to Actigraph, not significant (good quality)	0	0
Heesch et al. (2011) (IPAQ-SF)	Community dwelling elderly adults, mean age 73 years	N= 41	0	0	0	0	Content validity showed that definitions were confusing and recall was difficult (good quality)
Hurtig-Wennlof et al. (2010) (IPAQ-SF)	General population, mean age 71-74 years	N= 54	0	0	Kappa agreement with accelerometer,	0	0

					Kappa= 0.45 (good quality)		
Mader et al. (2006) (IPAQ-SF)	General population, mean age 45-57 years	N= 35	Spearmans rank=0.54 (poor quality)	0	Correlation to accelerometer, r=0.30-0.33 (poor quality)	0	0
Tomioka et al. (2011) (IPAQ-SF)	Adults part of a cohort study, mean age 69-77 years	N= 349	ICC=0.5-0.65 (excellent quality)	0	Correlation to accelerometer r=0.38-0.56 (good quality)	0	0
Blikman et al. (2013) (IPAQ-SF & IPAQ-LF)	Adults with TKA or THA due to OA, mean age 72 years	N=44	IPAQ-SF=ICC 0.51 IPAQ-LF= ICC 0.65 (fair quality)	0	Correlation to accelerometer IPAQ-SF (r=0.29) IPAQ-LF (r=0.43) (fair quality)	0	0
Cerin et al. (2012) IPAQ-SF	Adults aged 65 years or over, mean age not reported	N=94	ICC=0.86 (good quality)	0	Correlation to accelerometer total scores r=0.39 (good quality)	0	0
Dyrstad et al. (2013) IPAQ-SF	Community dwelling adults, mean age 48 years	N=1751	0	0	Correlation to accelerometer total scores r=0.33 (excellent quality)	0	0
Hansen et al. (2014) IPAQ-LF	Healthy adults, mean age 49.0 years	N=121	ICC=0.58 (good quality)	0	Correlation to accelerometer/heart rate monitor, males (r=0.46), females (r=0.25) (good quality)	0	0
Milanovic et al. (2014) IPAQ-LF	Community adults, mean age of 67 years	N=660	ICC=0.74 (good quality)	0	0	0	0

Sebastião et al. (2012) IPAQ-LF	Community adults, mean age 46-49 years	N=1572	0	0	0	0	Overestimations in household and work activities (fair quality)
Tran et al. (2013) IPAQ-LF	Community dwelling adults mean age 66 years	N=150	ICC=0.91 (poor quality)	0	Correlation to pedometer, r=0.20 (poor quality)	Correlation to activity logbook, r=0.46 (fair quality)	0
Van Holle et al. (2015) IPAQ-LF	Community dwelling adults	N=434 validity N=29 reliablitity	ICC=0.63 (poor quality)	0	Correlation to cut-off scores in accelerometer Copeland, 0.36 Freedson, 0.40 (poor quality)	0	0

Key: Excellent quality, good quality, fair quality refers to the methodological quality of the study measurement property assessed using the COSMIN checklist

(Mokkink et al., 2010), 0 = no evaluation of the measurement property within the study.

Physical Activity Scales for the Elderly (PASE)

The PASE questionnaire is a multidimensional PA instrument that was developed for elderly adults aged 65 years and above (Washburn et al., 1993). The PASE has 32 items covering six different dimensions and has questions on PA in leisure, occupational and household activities. The PASE was developed through reviewing the literature of self-report PA instruments, from which dimensions of PA for inclusion were proposed to two leading consultants in PA (Washburn et al., 1993). After a consensus meeting, drafts of the PASE were tested in a small sample of adults aged 65 years and over (Washburn et al., 1993).

The PASE can be administered either by self-completed questionnaire or by telephone administration. The PASE asks participants to recall activities over the previous seven days. The PASE gives an overall score based on the weighting of items depending on the intensity of exercise e.g. light, moderate or vigorous, calculated via objective activity monitoring and exercise diaries (Washburn et al., 1993). Scoring of the PASE can be quantified in a single score for responders' PA which ranges from 0-400 (Washburn et al., 1993), this is not interpretable for evaluating the average energy expenditure in a week. Table 8.22 summarises the qualitative attributes of the PASE (Washburn et al., 1993).

Table 8.22 Qualitative attributes of PASE

Qualitative Attribute	Definition
Construct	Time spent participating in PA
Setting	Long version includes; transport, housework activities, leisure time sports and time spent sitting. Short version does not use settings
Recall period	Leisure activities, occupational activities and household activities
Purpose	Research to assess PA in elderly adults
Target population	Elderly adults, aged 65 years and over
Justification	None of the generic measures of PA are appropriate for elderly adults
Format	32 items within the six different domains
Interpretability	Scores given as a total score, total score not interpretable in a meaningful way
Ease of use	Questions are easy to fill out with full instruction, short recall period, 32 items is a high number and so take some time.

(1) Reliability

Two studies (Martin et al., 1999, Svege et al., 2012) assessed the measurement properties of the PASE questionnaire in OA or joint pain populations. One of these examined the reliability, measurement error and construct validity of the PASE in a Norwegian population with hip pain in the last three months and evidence of radiographic OA in the hip (Svege et al., 2012). Findings showed the PASE to be reliable in test re-test reliability (ICC=0.77), the sample size for this analysis was small (n=33). This study also calculated the standard error of measurement (SEM) which was 31 and the minimal detectable change (MDC) which was 87.

In a community dwelling population of adults aged 45 years and over there were nine studies on the measurement properties of the PASE questionnaire (Colbert et

al., 2011, Dinger et al., 2004; Hagiwara et al., 2008; Harada et al., 2001; Moore et al., 2008; Schuit et al., 1997; Washburn & Ficker, 1999; Washburn et al., 1993; Washburn et al., 1999). Four evaluated the reliability of the PASE in adult populations with Dinger et al. (2004) reporting reliability in the PASE when administered by interviews (ICC=0.91, n=56). Three studies tested reliability of self-administered forms of the PASE: Washburn et al. (1993) reported reliability ICC=0.68 & 0.84 (n=244); Hagiwara et al. (2008) found that the PASE had an ICC=0.65 (n=257); Colbert et al. (2011) tested the reliability of a modified version of the PASE, although the nature of the modification was not explicit, the ICC was r=0.6, (n=55).

(2) Criterion validity

Two studies measured criterion validity by comparing PASE scores in elderly adults with those of a considered gold standard measurement of energy expenditure: DLW. One study found no significant correlation (n=56) (Colbert et al., 2011) and the other study had (n=21) found moderate correlations with DLW (r=0.58) (Schuit et al., 1997).

(3) Construct validity

The construct validity of the PASE has been assessed with accelerometers (Actigraph GT1M) and IPAQ scores in the same study of adults with hip OA (Svege et al., 2012). The sample size in both of these studies' analyses were small, n=36 and n=26 (Svege et al., 2012). A weak correlation between the PASE and Actigraph scores (r=0.3) was found, and a strong correlation was found with the IPAQ (r=0.61) (Svege et al., 2012).

Five studies have assessed the construct validity of the PASE compared with objective accelerometers in community dwelling adults (Washburn & Flicker, 1999; Harada et al., 2001; Colbert et al., 2003; Dinger et al., 2004; Hagiwara et al., 2008). Two of these studies compared total PASE scores to those of total PA counts measured using Actigraph PA monitors (Colbert et al., 2003, Dinger et al., 2004). Dinger et al. (2004) found a moderate association of PASE score with Actigraph (r=0.43, n=56) and Colbert et al. (2011) found a moderate association between the modified version of the PASE to the Actigraph (r=0.36, n=56). Three other studies had assessed construct validity in the PASE: Washburn & Flicker (1999) associated total PASE scores with total acceleration measured using the Computer Science and Application portable accelerometer and found in a small sample (n=20) a moderate association (r=0.49). Hagiwara et al. (2008) correlated total PASE score with Life Corder accelerometer measure of energy expenditure, finding a weak correlation with a Spearman's rank correlation (r=0.16) in a large study sample (n=325). Finally, Harada et al. (2001) found a moderate association with the Mini-Log activity monitor (r=0.59), when Mini-Log was measured at the ankle and at the waist (r=0.52) in a large study sample (n=87).

Two studies assessed the PASE by correlating scores with measures of physical function and self-report physical function (Pre-EPIC questionnaire, self-report pain intensity and The Continuous Scale Physical Functional Performance 10-item Test (CS-PFP10)) (Martin et al., 1999, Moore et al., 2008). In one study participants had chronic knee pain and were from a large cohort (n=471) (Martin et al., 1999), here the PASE was moderately associated with self-report physical function, using the Pre-EPIC questionnaire (r=0.35), 6-minute walking test (r=0.35) and knee strength (r=0.41) (Martin et al., 1999); there were non-significant correlations to

self-report pain intensities with the PASE score (Martin et al., 1999). Another study included participants from general populations aged 50 years and over found PASE was moderately correlated (r=0.4) with the CS-PFP10, in a sample of 54 participants (Moore et al., 2008). Table 8.23 summarises the measurement properties of the PASE, methodological quality was reported in brackets in each measurement property for each study.

(4) Updated search

Four new articles were identified in the updated search that evaluated measurement properties of the PASE. Bolszak et al. (2014) evaluated the construct validity, reliability and measurement error of the PASE in patients with a total knee replacement, the study sample was stratified by gender and mean age was 70 ± 6 years. In males (n=25), reliability of the PASE was found to be acceptable (ICC=0.77) and lower (ICC=0.58) in females (n=25). SEM was 32% for males and 35% for females; SDC was reported as 89% for males and 97% for females. There was a large difference between males and females in correlations between the PASE and the Actigraph GT3X. In males, there was a moderate correlation between PASE and the Actigraph GT3X (r=0.45) and a weak correlation in females (r=0.06). Females in this sample were older and had a higher BMI compared with the males, suggesting there may be other factors that have contributed to the difference in males and females.

A similar study was also conducted in patients with a total hip replacement (Casartelli et al., 2015). The measurement properties were evaluated in those that had recently undergone a hip replacement (within 2 months to 7 months, n=25) and those that had a hip replacement between 7 and 12 months ago (n=25). In the

total sample mean age was 68.3±5.9 years (n=50). Reliability of the PASE was reported as acceptable (ICC=0.77) with a SEM of 23% and SDC of 63.8%. The group that had a more recent hip replacement had better reliability and measurement error; however it was not significantly different from the other hip replacement group. In the evaluation of construct validity, a low correlation was found in the total sample between the PASE total scores and the Actigraph GT3X (r=0.27). There was no significant difference between the two groups in construct validity of the PASE.

Ngai et al. (2014) also evaluated the construct validity and reliability of a Chinese translated version of the PASE. The translation of the PASE was described as independently translated into Chinese, then backwards translated into English. The translated version was then assessed in an expert panel committee, with example activities modified to reflect usual Chinese physical activities. The final version was then agreed. The Chinese version of the PASE was tested in 90 adults with a mean age of 77.7±7.7 years. Reliability of the PASE was found to be acceptable (ICC=0.81). Validity of the Chinese version PASE was evaluated using comparisons to physical functioning assessments including the SF-36 (Lam et al., 1998), grip strength, quadriceps strength a balance test (single leg stance), 10 minute walking test and a 5 times sit-to-stand test (Ngai et al., 2014)

There were significant correlations between the PASE and six of the domains of the SF-36 (physical functioning, role physical, bodily pain, general health, vitality and social role). Grip strength, the balance test, the 5 times sit-to-stand and the 10 minute walking test were all significantly correlated to the Chinese version of the PASE, although quadriceps strength was not.

Another study also evaluated another Chinese version of the PASE used in a Chinese population in Canada (Vaughn & Miller, 2013). This version of the Chinese PASE appears to be a separate translated version to the previous study (Ngai et al., 2014). The translated version of the PASE was forward and backward translated by two independent translators, with disagreement then discussed in a meeting between the two translators and a mediator. Construct validity, reliability and measurement error were evaluated in this Chinese version of the PASE in 73 adults (71% females) with a mean age of 76.0 ± 49.1.

Construct validity was tested by evaluating the strength of correlation between the PASE and a number of physical functioning assessments. The physical functioning assessments included self-report measures such as the Activities-specific Balance Confidence Scale (Doble & Fisher, 1998) and Older American Resource and Services – Activities of Daily Living Scale (Powell & Myers, 1995)). The scores in PASE were also correlated to the Timed-Up-and-Go performance based measure (Podsiadlo & Richardson, 1991). Hypothesis testing was used in the evaluation of construct validity. It was hypothesised that the PASE would positively correlate to all the other measures at a magnitude of r=0.5 or greater. The PASE correlated with the Timed-Up-and-Go test (r=0.52); the Older American Resource and Services – Activities of Daily Living Scale (r=0.56) and the Activities-specific Balance Confidence Scale score (r=0.62). Reliability was evaluated using a two week test-retest where the ICC was found to be high (ICC=0.79). Measurement error was small (SEM=22.77 and SDC=63.11) and the 95% limits of agreement for the PASE were 70.9 and -88.5.

Table 8.23 Summary of the PASE measurement properties

Article	Population	Sample size	Reliability	Criterion validity	Construct validity using objective measure	Construct validity using physical function
Martin et al. (1999)	Knee pain patients from trial study, means age 71 years	N=471	0	0	0	Correlations with pain scores, not significant (poor quality)
Svege et al. (2012)	Hip OA patients, mean age 61 years	N=40 (n=33 reliability study)	ICC=0.77, (poor quality)	0	Correlation to Actigraph, r=0.30 (poor quality)	0
Colbert et al. (2011)	General population, mean age 75 years	N= 56	ICC= 0.60 (good)	Correlation with DLW, not significant (excellent quality)	Correlation with Actigraph, 0.36 (good quality)	0
Dinger et al. (2004)	Adults from a nutritional programme, mean age 76 years	N=56	ICC=0.91 (good quality)	0	Correlation with Actigraph, 0.43 (good quality)	0
Hagiwara et al. (2008)	General population, mean age 73 years	N= 325	ICC=0.65 (good quality)	0	Correlation with accelerometers, Spearmans Rank, 0.16 (fair quality)	0
Harada et al. (2001)	Community centre adults, mean age 75 years	N= 87	ICC=0.75 (good quality)	0	Correlation with accelerometers, 0.59 for ankle monitor, 0.52 for hip monitor (good quality)	Correlation to physical performance, 0.30 (fair quality)

Moore et al. (2008)	General population over 50, mean age 70	N=54	0	0	0	Correlation with physical function performance score, r=0.54 (fair quality)
Schuit et al. (1997)	Population from an intervention study, mean age 69-70 years	N=21	0	Correlation with DLW, 0.58 (poor quality)	0	0
Washburn & Ficker (1999)	Healthy volunteers to study, mean age 72 years	N= 20	0	0	Correlation to accelerometer r=0.49 (poor quality)	0
Washburn et al. (1993)	Adults, mean age 73 years	N= 396	ICC= 0.68-0.84 (good quality)	0	0	Correlation with perceived health status , r=-0.34 (fair quality)
Bolszak et al. (2014)	Adults after knee replacement, mean age 70 years	N=25 males N=25 females	Males ICC= 0.77 Females ICC=0.58 (fair quality)	0	Correlation to accelerometer males (r=0.45), females (r=0.06) (fair quality)	0
Casartelli et al. (2015)	Adults after hip replacement, mean age 68 years	N=50	ICC=0.77 (good quality)	0	Correlation to accelerometer r=0.27 (good quality)	0
Ngai et al. (2014)	Community Chinese adults, mean age 77 years	N=90	ICC=0.81 (fair quality)	0	0	Significant relationship with self-report and performance based measures (fair quality)

Vaugh & Miller (2013)	Community dwelling adults, mean age 76 years	N=73	ICC= 0.79 (good quality)	0	0	R=>0.5 for self-report and perforamance based measures (fair quality)

Key: Excellent quality, good quality, fair quality refers to the methodological quality of the study measurement property assessed using the COSMIN checklist (Mokkink et al., 2010), 0 = no evaluation of the measurement property within the study.

The short questionnaire to assess health enhancing physical activity (SQUASH)

The SQUASH aims to measure the habitual activities in a normal week over the past few months (de Hollander et al., 2012). The SQUASH assesses time spent and frequency of light, moderate and vigorous activities and is scored and compared with guidelines on recommended levels of PA (de Hollander et al., 2012). Table 8.24 provides a summary of the qualitative attributes of the SQUASH questionnaire (de Hollander et al., 2012).

Table 8.24 Qualitative attributes of the SQUASH questionnaire.

Qualitative Attribute	Definition
Construct	Habitual activities
Setting	Leisure activities, travelling activities, household activities, activities at work
Recall period	Normal week over past few months
Purpose	A self-report measure with comparable scores to recommendations of levels of physical activities for health benefits; to be used in epidemiology studies
Target population	All adult populations
Justification	Required a measurement where scores were interpretable to quantify weekly PA levels
Format	Eleven items asking questions on PA in different settings
Interpretability	Scores can be classified for recommended PA levels
Ease of use	Very short, simple to complete

(1) Construct validity

Only one study was found in the systematic review that examined the measurement properties of the SQUASH (de Hollander et al., 2012), De Hollander et al. (2012) studied community dwelling adults (n=187), aged between 28-60 years with a mean age of 57 (±11) years. Construct validity was tested by measuring agreement between the SQUASH categories of PA with the Actiheart monitor, a combination heart rate monitor and an accelerometer. The SQUASH showed a 37.4% agreement to the American College Sport Medicine (ACSM) guidelines (Haskell et al., 2007) for activity and 20.9% maximum disagreement with the Actiheart measure. Further analysis showed that when stratified for age, agreement was lowest and disagreement was highest in the subsample aged 55 years and above compared to younger ages (de Hollander et al., 2012). Table 8.25 summarises the measurement properties of the SQUASH, methodological quality was reported in brackets in each measurement property for each study.

Table 8.25 Summary of the SQUASH measurement properties

Article	Population	Sample size	Reliability	Criterion validity	Construct validity using objective measure	Construct validity using physical function
De Hollander et al. (2012)	Adult population, mean age 57 years	N=187	0	0	agreement for combined guidelines with heart monitors, agreement= 78.6% age specific (fair quality)	0

Key: Excellent quality, good quality, fair quality refers to the methodological quality of the study measurement property assessed using the COSMIN checklist (Mokkink et al., 2010), 0 = no evaluation of the measurement property within the study.

The short telephone activity rating (STAR)

The STAR was developed as a quick-to-administer telephone interview tool to assess levels of PA in adult populations (Matthews et al., 2005). Two versions of the STAR are available, the open response question version, which allows responders to give open answers and the closed response question version, which give responders a set number of choices for each question. Both versions contain just three questions on moderate and vigorous PA, and were developed based on already existing self-report PA questionnaires (Matthews et al., 2005). Responses can be scored into three levels of PA: inactive, somewhat active and active (Matthews et al., 2005). Table 8.26 provides a summary of the qualitative attributes of the STAR questionnaire (Matthews et al., 2005).

Table 8.26 Qualitative attributes of the STAR questionnaire.

Qualitative Attribute	Definition
Construct	Classification of PA in moderate and vigorous levels of PA
Setting	All PA
Recall period	Last 7 days
Purpose	A telephone administered short instrument to classify individuals in different levels of PA
Target population	All adult populations
Justification	A need for a quick-to-complete measure of PA over the telephone, scores can classify responders in different levels of PA
Format	3 items, two versions available; open responses and closed responses
Interpretability	Responders can be classified into different levels of PA
Ease of use	Very quick to administer

(1) Reliability

Only one study examined the measurement properties of the STAR (Matthews et al., 2005). Reliability was evaluated in a sample of community dwelling adults (n=108) with an average age of 46 years (standard deviation was not reported). Reliability was measured by evaluating agreement in scores three days apart. The open version of the STAR showed moderate agreement in repeated measures (Kappa= 0.57) and the close version showed higher agreement in repeated measures (Kappa=0.76).

(2) Construct validity

Construct validity was assessed for both versions of the STAR, participant's PA score for the previous 24 hours and accelerometer (Actigraph) measured activity levels were evaluated for agreement. Moderate agreement was found in both the open version (Kappa= 0.43) and closed version (Kappa= 0.36) to a 24 hour PA recall. Weak agreement was found in the open version (Kappa= 0.14) and closed version (Kappa= 0.15) when compared with the Actigraph measurement of PA levels (Matthews et al., 2005). Table 8.27 summarises the measurement properties of the STAR, methodological quality was reported in brackets in each measurement property for each study.

Table 8.27 Summary of the STAR measurement properties

Article	Population	Sample size	Reliability	Criterion validity	Construct validity using objective measure	Construct validity using physical function
Matthews et al. (2005)	General adult population, mean age 46 years	N=108	Open version Kappa= 0.57, closed version Kappa=0.76 (excellent quality)	0	Correlation with Actigraph open version Kappa 0.14, closed version Kappa 0.15 (good quality)	0

Key: Excellent quality, good quality, fair quality refers to the methodological quality of the study measurement property assessed using the COSMIN checklist (Mokkink et al., 2010), 0 = no evaluation of the measurement property within the study.

Zutphen questionnaire

The Zutphen questionnaire is a self-report PA measure that was developed for assessing the levels of PA in a large longitudinal epidemiological study in the Netherlands, the Zutphen study (Harris et al., 2008). The Zutphen questionnaire consists of 17-items on different dimensions of activities conducted over the previous seven days and in some items over an average week; it was adapted from a previous questionnaire developed to assess levels of activities in retired males (Caspersen et al., 1991). The scoring of the Zutphen questionnaire consists of scoring minutes spent per week participating in certain activities and estimating the intensity of the activity, from this, frequency, duration and intensity of exercise can be calculated and individual scores can be interpreted according to PA recommendations (Caspersen et al., 1991). Table 8.28 provides a summary of the qualitative attributes of the Zutphen questionnaire (Caspersen et al., 1991).

Table 8.28 Qualitative attributes of the Zutphen questionnaire

Qualitative Attribute	Definition
Construct	Daily physical activities
Setting	Leisure-time, walking, household activities, sporting activities and hobbies.
Recall period	7 days, although some items differ
Purpose	Used to assess levels of PA in a longitudinal study
Target population	Designed for a study in older male adults, but has been used in male and female adults since
Justification	Developed as an appropriate measure of PA over time for a longitudinal study
Format	17 items, open and closed questions
Interpretability	Total score given as energy expenditure,
Ease of use	Short with minimal requirements for completion

(1) Construct Validity

Only one study evaluated the measurement properties of the Zutphen questionnaire (Harris et al., 2009). The study was conducted in male and female adults aged 65 years and over. Scores for the Zutphen questionnaire were compared to activity counts measured by the Actigraph accelerometers and step count measured by a pedometer (Harris et al., 2009). The Zutphen questionnaire was found to be moderately correlated with both the Actigraph (r=0.34) and pedometer (r=0.35) in an adequately sized sample (n=234). Table 8.29 summarises the measurement properties of the Zutphen questionnaire, methodological quality was reported in brackets in each measurement property for each study.

Table 8.29 Summary of the Zutphen questionnaire measurement properties

Article	Population	Sample size	Reliability	Criterion validity	Construct validity using objective measure	Construct validity using physical function
Harris et al. (2009)	Community elderly population, mean age 74 years	N=234	0	0	Correlation with Actigraph, 0.34, and pedometer, 0.35 (good quality)	0

Key: Excellent quality, good quality, fair quality refers to the methodological quality of the study measurement property assessed using the COSMIN checklist (Mokkink et al., 2010), 0 = no evaluation of the measurement property within the study.

Single item scale questionnaires

One study was found that assessed the measurement properties of three self-report single scale items of PA: University of California Los Angeles Activity (UCLAA) scale, the Tegner scale and the Activity Rating Scale (ARS) (Naal et al., 2009). The UCLAA scale is a simple scale that ranges from 1 to 10 with 1 indicating no participation in PA and 10 representing regular participation in impact sports. The Tegner score is similar to the UCLAA score, with a scale ranging from 0, indicating no PA and being disabled to 10, which represents participating in elite level sporting activities. The ARS gives a total score of PA covering different types of activities (running, changing direction while running, decelerating and pivoting). Table 8.30 provides a summary of all of the single item scale questionnaires qualitative attributes (Naal et al., 2009).

Table 8.30 Qualitative attributes of the single item scale questionnaires

Qualitative Attribute	Definition
Construct	Physical activities
Setting	All PA
Recall period	Last 7 days
Purpose	To assess level of PA in one item
Target population	All adult populations
Justification	No valid single item measure of PA
Format	One item with a scale of possible answers to the item
Interpretability	Each value on the scale identifies individuals at an interpretable level of PA
Ease of use	Only one item, very quick and easy, minimal burden for responders

Reliability

Naal et al. (2009) evaluated the measurement properties of these single item-scales, in patients undergoing total hip replacement (THR) (n=105) and total knee replacement (TKR) (n=100); test retest reliability was tested in a smaller sub sample (n=43 THR, n=36 TKR). The UCLAA scale was found to be reliable in the THR and TKR groups, with Kappa = 0.80, Kappa = 0.86 respectively. Reliability in the ARS and Tegner scales were lower in the THR group (ARS Kappa = 0.65, Tegner Kappa = 0.54), nevertheless, both showed reliability in the TKR group (ARS Kappa = 0.88, Tegner Kappa = 0.84). The study also found none of the scales had a floor or ceiling effect, which can occur in single scale items (Naal et al., 2009). Table 8.31 summarises the measurement properties of the single item scale questionnaires, methodological quality was reported in brackets in each measurement property for each study.

Table 8.31 Summary of the single scale item questionnaires measurement properties

Article	Population	Sample size	Reliability	Criterion validity	Construct validity using objective measure	Construct validity using physical function
Naal et al. (2009)	Total knee + hip replacement patients, mean age 63-66 years	N=43 THR, N=36 TKR	ICC range for all single scale items ICC=0.54-0.88 (fair quality)	0	0	0

Key: Excellent quality, good quality, fair quality refers to the methodological quality of the study measurement property assessed using the COSMIN checklist (Mokkink et al., 2010), 0 = no evaluation of the measurement property within the study.

8.5.5 Grading of instruments' measurement properties

Table 8.32 and 8.33 demonstrate the grading system described in section 8.4.2 developed for this systematic review used to indicate the level of strength in measurement properties for each of the instruments included in stage B.

Table 8.32 Grading of measurement properties in instruments in adults with joint pain or OA

Instrument	Reliability	Criterion validity	Construct validity using objective measure	Construct validity using physical function	Internal consistency	Content validity	Structural validity	Responsive-ness
AAS	0	0	0	0	0	0	0	0
ARS*	+	0	0	-	0	0	0	0
Beacke	++	0	0	--	0	0	0	0
Modified Beacke	0	0	0	0	0	0	0	0
HAP	+++	0	0	--	0	0	0	0
IPAQ	+++	0	?	0	0	0	0	0
IPEQ	0	0	0	0	0	0	0	0
PASE	++	0	--	0	0	0	0	0
SQUASH questionnaire	0	0	0	0	0	0	0	0
STAR questionnaire	0	0	0	0	0	0	0	0
Tegner*	+	0	0	-	0	0	0	0
UCLAA*	+	0	0	-	0	0	0	0
Zutphen Questionnaire	0	0	0	0	0	0	0	0

Key: ? indicate unclear findings due to study quality, ± indicates conflicting findings , Strength of the evidence was given based on quality of articles assessed by the COSMIN checklist and Table 8.6. Instruments were given a positive, negative or zero score for the corresponding measurement property based on criteria in Table 8.7.

Table 8.33 Grading of measurement properties in instruments in community dwelling adults aged 45 years and over

Instrument	Reliability	Criterion validity	Construct validity using objective measure	Construct validity using physical function	Internal consistency	Content validity	Structural validity	Responsive-ness
AAS	++	0	---	--	0	---	0	0
ARS*	0	0	0	0	0	0	0	0
Beacke	0	0	0	0	0	0	0	0
Modified Beacke	?	?	?	-	0	0	0	0
HAP	0	0	--	0	0	0	0	0
IPAQ	±	0	---	-	0	---	0	0
IPEQ	++	0	---	0	0	0	0	+++
PASE	±	--	---	--	0	0	0	0
SQUASH questionnaire	0	0	+	0	0	0	0	0
STAR questionnaire	±	0	--	0	0	0	0	0
Tegner*	0	0	0	0	0	0	0	0
UCLAA*	0	0	0	0	0	0	0	0
Zutphen Questionnaire	0	0	--	0	0	0	0	0

Key: ? indicate unclear findings due to study quality, ± indicates conflicting findings , Strength of the evidence was given based on quality of articles assessed by the COSMIN checklist and Table 8.6. Instruments were given a positive, negative or zero score for the corresponding measurement property based on criteria in Table 8.7.

From Tables 8.32 and 8.33, it can be seen that some instruments had positive evidence for some of the measurement properties. Every instrument had missing evidence for some of its measurement properties, with no instrument considered to be a fully valid and appropriate measure of self-report PA. There was no measurement property evidence for the AAS, modified Baecke, IPEQ, SQUASH, STAR and Zutphen in adults with joint pain or OA (Table 8.32). In both populations, there was no evidence for internal consistency or structural validity for any of the instruments. Only one instrument, the IPEQ, had evaluated responsiveness (Table 8.33). None of the included studies evaluated content validity in adults with joint pain or OA.

The IPAQ (both formats) and PASE had positive evidence for reliability in adults with joint pain or OA (Table 8.32). This evidence was a mixture of positive and negative evidence for reliability in community dwelling adults aged 45 years and over (Table 8.33). In two studies evaluating reliability, methodological quality was poor to fair with sample sizes below 50 participants. Sample sizes below 50 are considered too small for evaluating measurement properties (Terwee et al., 2007). Studies that evaluated criterion validity were small with less than 50 participants (Table 8.33). Both the IPAQ and PASE showed only low to moderate correlations with more objective measures of PA. The PASE and modified Baecke were the only instruments to have evidence for criterion validity in community dwelling

adults aged 45 and over, although the relationship between the two instruments and DLW was statistically insignificant (Table 8.33).

Other instruments with positive findings in at least one measurement domain included the single-scale items and the HAP. The single-scale items and HAP had positive evidence for reliability in adults with joint pain or OA (Table 8.32).

8.6 Discussion

This section discusses the main findings of this study in the overall aim of this chapter's study as outlined in section 8.2 and the context of each of the objectives outlined in section 8.3. This section makes comparisons of the findings of this study in relation to other relevant research. The strengths and limitations of this study and the possible future research based on this study are discussed in Chapter ten of this thesis.

This chapter has described the methods, results and findings of a systematic review of self-report instruments of PA for the target population. The systematic review builds on prior knowledge of the measurement properties of self-report PA instruments in the target population from a previous systematic review (Terwee et al., 2011). This was achieved by identifying instruments used within OA and joint pain research for assessing levels of PA and then evaluating the measurement properties of those instruments. Measurement properties of the identified instruments were evaluated in two groups: the target population and community dwelling adults 45 years and over. This allowed for a larger pool of evidence to be collected and assessed than in previous systematic review on the OA population

(Terwee et al., 2011), in adult populations aged 18-65 years (van Poppel et al., 2010) and in elderly populations (Forsen et al., 2010).

8.6.1 Stage A: Main findings

The study aimed to identify which self-report measures of PA were most appropriate for assessing levels of PA in an OA and joint pain population for use in primary care research. In Stage A 23 different self-report instruments were eligible, Stage B measurement properties of 13 of those 23 instruments were eligible for further analysis.

Stage A of this systematic review identified 23 self-report instruments of PA that had been previously used in the target population. Stage A also showed that the most frequently used self-report PA instruments in joint pain or OA research are the PASE (sixteen studies), IPAQ (both SF and LF) (ten studies), UCLAA scale (nine studies) and ARS (four studies). In twenty five studies identified in Stage A, self-report levels of PA were also measured using questions that had been derived for those studies. Using such questions as an outcome measure leads to risk of reporting bias as well as bias in recall which are difficult to establish as the questions used cannot be evaluated for validity or reliability (De Vet et al., 2011). Using validated instruments to assess levels of PA minimises reporting bias as the measurement properties of those instruments can be assessed (Mokkink et al., 2010).

8.6.2 Stage B: Main findings

Instruments that were identified in Stage A but did not have any measurement properties studies retrieved in Stage B included: the DAQ (Terwee et al., 2011,

Wollmerstdt et al., 2010), AAP (Foley et al., 2003), MLT-PAQ (Weller et al., 2006), LASA (Verweij et al., 2009), PASIPD (Groot et al., 2008) and LEAS (Terwee et al., 2011). In the updated search a further three instruments were identified in Stage A, none of these instruments identified had articles included in Stage B that evlauted their measurement properties. These instruments were: the Historical Leisure Activity Questionnaire (Jones et al., 2012), The Physical Activity Scale (Holm et al., 2014) and Yale Physical Activity Survey (Chang et al., 2014).

From this systematic review no recommendation can be made on which instrument is most appropriate for measuring self-report PA in joint pain or OA populations. It has, provided a summary and appraisal of the current evidence highlighting gaps and limitations of the identified instruments.

When examining the measurement properties of all instruments in Stage B, similarities across the instruments were apparent. The IPAQ, PASE, Modified Baecke, HAP, AAS demonstrated reliability. The Zutphen questionnaire, Baecke, STAR, single scale items and IPEQ had evidence of reliability in a single study. This suggests that most self-report instruments are reliable in measuring levels of PA in test re-test evaluations. The single scales items, HAP, IPAQ and PASE were the only instruments with evidence of reliability in the target population, other instruments showed reliability in community dwelling adult populations.

The findings for construct validity and criterion validity in Stage B showed correlations that were no higher than 0.70 in the target population and community adult populations. The small number of studies found in Stage B which compared the self-report PA questionnaire with DLW measurement of PA showed mixed findings (Colbert et al., 2011, Hertogh et al., 2008, Schuit et al., 1997).

Notably, only two studies evaluated content validity, those were for the AAS and IPAQ (Heesch et al., 2011; 2014). No articles evaluated face validity, internal consistency, structural validity or cross-cultural validity in any of the instruments in any population. Some instruments that had been translated into languages other than English had their translation described, although data on cross-cultural validity were never reported. There was no content validity for PA measures in adults with joint pain or OA, only in community-dwelling adults (Heesch et al., 2011; 2014).

None of the self-report instruments evaluated in Stage B appear to be fully validated for measuring of PA in OA or joint pain populations, or community dwelling adult populations. The PASE and the IPAQ rated most highly for assessing PA in terms of frequency, duration and intensity of PA. Both the PASE and the IPAQ had limitations in their measurement properties. In the IPAQ, the only one study evaluated IPAQ-LF, the other 8 studies exclusively evaluated the IPAQ-SF. The IPAQ showing some positive and negative reliability and both measures showed moderate associations with objective measures of PA. If these instruments are used in the target population, both are reliable and have some associations with other measures of PA although consideration should be given to the age of the population of interest. The IPAQ seems to score high relative reliability with ICCs in studies in adults below 65 years and the PASE was designed for elderly adults aged 65 years and over.

Other instruments evaluated in Stage B had more limited evidence for their reliability and validity but may still be of interest for use. In situations where investigators wish to measure PA levels and functional capacity, the HAP may be

an appropriate measure. The HAP gives two scores: the average level of PA and the maximum intensity of activities an individual is able to achieve, which is an indicator of functional capacity. The limitations of the HAP include a small evidence base in the target population with no studies in adults aged 45 years and over. The HAP is a lengthy instrument with 94 items making it time-consuming to complete (Bennell et al., 2004).

The STAR is an instrument (Matthews et al., 2005) that has not been evaluated for measurement properties in OA or joint pain populations and the systematic review only identified the STAR's development article, which had some evidence for the measurement properties in adult populations 45 years and over. The STAR contains just three items and the score of the STAR can be calculated to fit with recommendations on moderate and vigorous intensities of activity. This makes the STAR an attractive short screening tool to quickly assess the level of PA an individual is achieving; it has previously been used to measure PA levels in a large cohort study of knee pain (Holden et al., 2013). The STAR was developed as a telephone administered self-report measure, so would need evaluating if it was to be used in another format, such as a face to face interview.

No study found in Stage B examined the responsiveness of any of the instruments identified for either OA or joint pain, or community dwelling adult populations. Responsiveness is important as a PA instrument needs to be able to detect changes in levels in PA so it can be used in trials as an outcome of an intervention to change PA levels (De Vet et al., 2011). Similarly a measure that was responsive to change would be of use in longitudinal studies that are, for example, exploring whether increases in pain are linked to reduced levels of PA.

8.6.3 Similarities and differences from Terwee et al. (2011)

In 2011, Terwee and colleagues conducted a similar systematic review, which also described and evaluated measurement properties of self-reporting PA instruments in adults with OA, but focussing on the hip or knee specifically. The methodological approaches of the two systematic reviews differed. That presented in this thesis involved conducting a two-stage systematic review, while Terwee et al. (2011) used a one-stage approach, identifying studies that evaluated measurement properties of PA instruments in their target population. The advantage of the two-stage approach was that it allowed identification of all instruments that have been previously used to measure self-reporting PA in the target population, and then described which instruments did or did not have their measurement properties evaluated.

The included populations in both systematic reviews also differed. For this thesis, adults aged 45 years and over with joint pain or OA in the hip, knee, hand or foot, and a general population of community-dwelling adults aged 45 and over without joint pain, were included. As mentioned previously, the Terwee et al. (2011) review included studies in patients with a diagnosis of OA in the knee or hip only. While the knee and hip are common OA sites, the inclusion of hands and feet was decided in order to reflect current NICE guidelines (2008, 2014). The term joint pain was added, as adults aged 45 years and over with joint pain in the knee, hip, foot or hand is likely to have OA, or be at risk of OA (NICE, 2008, 2014). The use of the term joint pain led to the addition of one further study to the systematic review in this thesis, and used the PASE instrument (Martin et al., 1999).

For Stage B, synthesis of instrument measurement properties was conducted for adults aged 45 years and over with joint pain, and community-dwelling adults also of the same age range. Terwee et al. (2011) included only those with knee and hip OA. The addition of community-dwelling adults aged 45 and over was chosen so that the systematic review could provide wider evidence of measurement properties in the PA instruments in adults aged 45 and over.

The similarities between Terwee's (2011) review and that of this thesis were in the approach to quality assessment and data extraction of studies, conducted using QAPAQ and COSMIN, resulting in comprehensive evaluations of study methods and data. The rationale for selecting QAPAQ and COSMIN in this thesis was that they were evidenced, standardised forms which were found to be feasible for the reviewing team to use.

Despite some differences in the methodological approach (two stage approach, and different population) between the two systematic reviews, the findings were similar. For all instruments, there was a lack of good quality evidence for instruments' measurement properties. Both systematic reviews also found that no studies had investigated responsiveness in any instrument.

8.6.4 Comparison of findings in other studies

The overall findings of Stage B were similar to other evaluations of self-report PA instruments in adult populations (Forsen et al., 2010, Helmerhorst et al., 2012, Van Poppel et al., 2010). These previous systematic reviews report positive scores for the reliability of PA instruments for adults (Forsen et al., 2010; Helmerhorst et al., 2012; Van Poppel et al., 2010). PA questionnaires have been shown to have weak

or moderate correlations with objective measures of PA in adults and elderly populations (Forsen et al., 2010; Helmerhorst et al., 2012; Van Poppel et al., 2010). This suggests that self-report instruments are reliable in adults aged 45 years and over with joint pain or OA but are not valid compared with objective measures of PA. None of the previous systematic reviews had identified any studies on the responsiveness of self-report instruments of PA in adults.

Helmerhorst et al. (2012) conducted a large systematic review comparing the reliability and validity of PA instruments that had been developed in the last 15 years compared to older instruments that were 15 years or older. Helmerhorst et al.'s (2012) systematic review found that there was no difference in reliability and validity between the newer instruments and older instruments. This is similar to the findings from Stage B of this systematic review; measurement properties of older instruments such as the Zutphen developed in 1985 (Caspersen et al., 1991) are not distinguishable in their measurement properties from those of more recent instruments such as the IPAQ developed in 2003 (Craig et al., 2003). This is surprising given that in later years more focus has been on quantifying levels of PA in many different populations and establishing the health benefits of PA (Blair et al., 2001). Despite this measurement properties of self-report instruments have remained stable.

8.7 Conclusion

This chapter identified self-reporting PA instruments used in OA or joint pain research, and evaluated their measurement properties. The systematic review identified a large number of instruments, but there was limited evidence of their measurement properties. The findings of this review demonstrate that all of the

instruments identified have similar, but modest levels of reliability and validity in their measurement properties. Two instruments that may be of further interest, based on their approach to measuring levels of PA, are the IPAQ-SF and PASE. Both of these instruments measure PA in terms of frequency, duration and intensity of activity, which is of importance when considering recommended PA levels. Although this systematic review did not show the IPAQ-SF and PASE as superior in their measurement properties, these two instruments still warrant further investigation. Stage A of this review showed that both instruments have been widely used in previous research in OA and joint pain populations. Further understanding of their measurement properties is required if these instruments are to be continually used within these populations. The IPAQ-SF appears to have greater evidence for its measurement properties compared to the IPAQ-LF, and also benefits from fewer items, therefore increasing its ease of use. Critically, responsiveness has not yet been reported for any self-report measure of PA identified for use in the target population. Therefore, a focus of future research is to further evaluate the measurement properties of the IPAQ-SF and PASE in adults aged 45 and over with joint pain, particularly for responsiveness, which has yet to be evaluated in these instruments for this population.

Chapter Nine: Assessment of measurement properties in the International Physical Activity Questionnaire Short Form (IPAQ-SF) and Physical Activity Scale for the Elderly (PASE) in an adult population with joint pain.

9.1 Introduction

This chapter contains the methods, results and discussion of the assessment of measurement properties of the IPAQ-SF and PASE. The findings of the systematic review in Chapter eight showed the IPAQ-SF and PASE to be promising instruments for assessing levels of PA in the target population.

There was not enough evidence to recommend either the IPAQ-SF or the PASE as appropriate instruments to measure self-report PA in the target population. Importantly for assessing PA changes over time, there is currently no evidence for the responsiveness of the IPAQ-SF or PASE in the target population. As a result, the IPAQ-SF and PASE have been selected for analysis of measurement properties. This chapter discusses the clinicmetric evaluation of the IPAQ-SF and PASE which includes evaluating the reliability, measurement error, construct validity and responsiveness.

9.2 Aim

The aim of this study was to evaluate the measurement properties of the International PA Questionnaire Short Form (IPAQ-SF) and the PA Scale for the Elderly (PASE) in adults aged 45 years and over with joint pain.

9.3 Objectives

The objectives of this evaluation of the IPAQ-SF and PASE were to:

3a. To compare the responses of individual items and total scores between the IPAQ-SF and PASE.

3b. To assess and compare the reliability and measurement error of the IPAQ-SF and the PASE.

3c. To assess and compare the construct validity of total score and sub-domains in the IPAQ-SF and PASE.

3d. To assess and compare the responsiveness of total score and sub-domains in the IPAQ-SF and PASE.

9.4 Statistical analysis

As described in Chapter six, data was taken from the MOSAICS baseline and three month follow up consultation questionnaire. Participants eligible for inclusion into this study were those that completed either the IPAQ-SF or PASE. Baseline characteristics of the responders to the IPAQ-SF or PASE were compared to non-responders to the IPAQ-SF or PASE at baseline to evaluate risk of non-response bias. Baseline characteristics included: participant demographics, pain symptoms, health status and uptake of PA as a treatment for joint pain symptoms.

9.4.1 Scoring of the IPAQ-SF and PASE

Both measures were scored according to the instruments' manual guidelines supplied by the authors. The scoring guideline manual for the short form IPAQ-SF

can be found on the official IPAQ-SF website (http://www.ipaq.ki.se/scoring.htm). Scoring was conducted for both the categorical output and continuous output of the IPAQ-SF. Full truncation of the IPAQ-SF data and exclusion of outliers that reported greater than 16 hours of total PA was conducted as recommended in the IPAQ-SF guideline manual. The MOSAICS's project paid for the licensed use of the PASE instrument and scoring protocol. No truncation of data was conducted on the PASE data as the responses to items in the PASE are closed, with only a range of responses possible. Outliers of the PASE were checked if individual's total score exceeded 400 as recommended in the scoring guideline.

9.4.2 Objective 3a

To compare the responses of individual items and total scores between the IPAQ-SF and PASE.

To describe the distribution of total scores of the IPAQ-SF and PASE at baseline, histograms were used, with distribution lines in all responders. Histograms were compared to see if the IPAQ-SF or PASE differed in the distribution of data in total scores.

9.4.3 Objective 3b

To assess and compare the reliability and measurement error of the IPAQ-SF and the PASE.

(1) Reliability

Reliability was measured in a sub-sample of responders who remained stable in their levels of PA. The inclusion in this component of the analysis was respondents that reported:

- Not trying muscle strengthening exercises or general fitness exercise in both the baseline and three month follow up questionnaires of the MOSAICS consultation questionnaire.

And

- Those reporting that they had tried muscle strengthening exercises or general fitness exercises in both the baseline and three month follow up.

These criteria identified a sub-sample of responders that were deemed stable in their PA behaviour, either not trying exercises or participating in the same exercises during the three month follow up period.

For the reliability analysis, only responders who had completed all items of the IPAQ-SF and PASE at baseline and three months were taken. This was selected so that reliability could be evaluated for both the total scores and all subdomains of the IPAQ-SF and PASE.

Descriptive statistics were used to describe the characteristics of those included in the reliability analysis: gender, age and IMD deprivation scores were reported. At three month follow up, global assessment of change in joint pain was also described for the reliability sub-sample. Changes from baseline to three month data were also reported for BMI, highest pain intensity, SF-12 physical component score and mental component scores, EQ-5D score, PASE score, IPAQ-SF $METS^{-1minute\text{-}1week}$ and IPAQ-SF category score. Changes in these scores were used as a proxy measure for predicting if the reliability sub-sample were likely to have remained stable in their level of PA. Comparisons of change for continuous scores were conducted using paired *t*-test, or paired sample Wilcoxon signed rank test on non-normally distributed data. McNemar's test was conducted for IPAQ-SF categorical data.

In the reliability analysis of the continuous IPAQ-SF and PASE data, some of the data was normally distributed and some were positively skewed, therefore two-way random intraclass correlations for absolute agreement (ICC) were used for the normally distributed data and for the skewed data of PASE and IPAQ-SF, Spearman rank coefficient was selected as data was self-reported and were scored at both time points. Single measure ICC were used as only one measurement of IPAQ-SF and PASE were conducted at baseline and at three month follow up.

Absolute agreement ICC assesses variance between scores due to error variance and the systematic error between scores in the two time points. For this study, this refers to the error in responders reporting PA at baseline and three months, and the systematic difference between reporting them. It is expected that the

systematic difference should be small, as there is no systematic rationale for a change in the error of respondents reporting PA at baseline and then at three month follow up.

A cut off value of ICC=0.7 in a significant ICC was selected as adequate reliability (De Vet et al., 2011). There was no recommended cut-off value for Spearman rank coefficient *r* value for adequate reliability and r=0.7 was selected to match the cut-off for the ICC.

Reliability was tested in separate domains within the IPAQ-SF and PASE. For the IPAQ-SF, the separate domains were: vigorous, moderate, walking and sitting activities. For the PASE, the separate domains were: sitting, walking, light, moderate, strenuous and muscle strengthening exercises. Reliability testing in separate domains of the IPAQ-SF and PASE allowed for identifying where reliability in the instruments was low. For dichotomous variables in the PASE's household activities and work related activities reliability was assessed using Kappa, with 95% confidence intervals, to evaluate agreement between baseline and three month follow up (De Vet et al., 2011). Reference values were taken from Landis & Koch (1997) for interpretation of Kappa reliability. Table 9.1 below summarises the interpretation of kappa values below.

Table 9.1 Interpretation of Kappa values

Kappa value ranges	Interpretation
1.0- 0.8	Almost perfect agreement
0.79-0.6	Substantial agreement
0.59-0.4	Moderate agreement
0.39-0.2	Fair agreement
0.19-0.0	Slight agreement

To assess reliability of the ordinal data in IPAQ-SF categorical output, quadratic weighted Cohen's Kappa was selected to assess agreement between baseline and three month follow up. A weighted Kappa allows for misclassification among adjacent categories to be less serious than longer ranged misclassification by adding a numerical weighting. Misclassifications occur when the categorical score at three months does not match that at baseline. Quadratic weighting rating for IPAQ-SF categories were: 0 weighting for same IPAQ-SF category, 1 weighting for adjacent IPAQ-SF categories and 4 weighting for 2 IPAQ-SF categories a part (Fleiss & Cohen, 1973). Quadratic weighted Cohen's Kappa were calculated using the website available at: http://faculty.vassar.edu/lowry/kappa.html (De Vet et al., 2011) to gain the kappa value and 95% CI. This website was used as the SPSS version 21 software package did not have the functionality to calculate a quadratic weighted Cohen's Kappa.

(2) Measurement error

For measurement error analysis, the same sub-sample of responders was used to those in the reliability analysis. Analysis of measurement error included reporting the standard error of measurement (SEM) and smallest detectable change (SDC) in the IPAQ-SF and PASE total scores where an ICC was calculated. As the ICC analysis was conducted for absolute agreement, SEM and SDC were also used in absolute agreement. SEM agreement was calculated as shown below (De Vet et al., 2011).

SEM agreement = $\sqrt{(\sigma^{2\circ} + \sigma^2 residual)}$

Where $\sigma^{2\circ}$ is the variance due to the systematic difference between time points and $\sigma^2 residual$ is the residual variance, such as random error due to a combination of responders and time points (De Vet et al., 2011).

The SEM agreement was calculated by hand using values for $\sigma^{2\circ}$ and $\sigma^2 residual$ taken from a VARCOMP output in SPSS software package. The syntax for the VARCOMP can be found in appendix 1.10. A VARCOMP is a feature in SPSS that analyses the variance component for general linear models with random effect. For this analysis only the $\sigma^{2\circ}$ and $\sigma^2 residual$ were extracted from the VARCOMP (De Vet et al., 2011).

SDC was calculated as ± 1.96 * $\sqrt{2}$ * SEM (De Vet et al., 2011). The value for SEM was taken from the SEM agreement calculation. SDC was again calculated by hand rather than coded on to SPSS for simplicity.

Bland and Altman plots were also used to display measurement error in the PASE and IPAQ-SF baseline and three month total scores. The Bland and Altman plots were created using mean scores of baseline and three month follow up in total score of the PASE and then IPAQ-SF, against difference in scores (three month follow up – baseline). Lower and upper 95% limits of agreement were then calculated using the mean difference in scores ± 1.96 * the SD of difference, in data where differences in scores were normally distributed (Bland & Altman, 1999). Where differences in scores were not normally distributed, an alternative method suggested by Bland & Altman (1999) for assessing upper and lower 95% limits of agreement in non-normally distributed data was used. This included ranking all the responders scores within a plot and then selecting the 5% top responders score and 5% bottom responders scores as value for the upper and

lower limits of agreement. The range of the upper and lower values can then be used to determine if the measurement error is acceptable in the given instrument (Bland & Altman, 1999). Selecting a cut off value as a limit of acceptable measurement error is minimal clinically important change/difference score. For example, if the measurement error is larger than the minimal clinically important change/difference, that instrument's measurement error is too large (De Vet et al., 2011). There is no minimal clinically important change/difference cut off value established for IPAQ-SF or PASE. In this study, no cut-off values were set for limit of agreement (Bland & Altman, 1999).

9.4.4 Objective 3c

To assess and compare the construct validity of total score and sub-domains in the IPAQ-SF and PASE.

To assess the construct validity, the full sample of respondents (n=310) to the baseline survey was taken. Construct validity of the PASE and IPAQ-SF was conducted using hypothesis testing. Four hypotheses for construct validity were created and tested. Box 9.1 summarises the four hypotheses that were tested to evaluate construct validity for objective 3c. These hypotheses were constructed following recommendations for creating hypothesis for construct validity (De Vet et al., 2011).

Box 9.1 Four hypotheses tested for construct validity in objective 3c

Hypotheses tested in objective 3c
The baseline total score of the PASE will be positively correlated with the baseline total IPAQ-SF score at a cut-off of r=0.5 or greater.
The baseline scores for time spent sitting, walking, moderately active and vigorously active subdomains of the IPAQ-SF and PASE will be positively correlated at r=0.5 or greater.
The baseline total scores of the IPAQ-SF and PASE will be positively correlated with physical component score of the SF-12 at r=0.2 or greater.
Respondents reporting more limitation in moderate activities in the SF-12 will reported significantly lower levels of PA in the PASE and IPAQ-SF compared to those reporting less limitation.

In the first construct validity hypothesis test, because the PASE and IPAQ-SF are measuring self-reported levels of PA, a correlation of r=0.5 or greater is expected. This was conducted using a Spearman's rank coefficient as data were not normally distributed. There are currently no guidelines for an appropriate value that can determine positive concurrent construct validity in health-related outcome measures. The cut off value of r=0.5 was thought to be appropriate, as this value is consistent with those in other examples of studies examining construct validity (Brouwer et al., 2007; De Vet et al., 2011). A value of r=0.5 would reflect both measures as closely correlated in measuring PA. This cut-off value was selected, rather than a high cut-off value, because the IPAQ-SF and PASE have been shown previously to contain large measurement errors in adults aged 45 years and over with OA (Svege et al., 2012; Blikman et al., 2013). A correlation of 0.5 represents a moderate correlation between two variables (Cohen, 1988). The justification for selecting 0.5 was to also reflect previous evidence for construct validity identified in Chapter Eight of this thesis. In that chapter, the IPAQ-SF and PASE were shown to correlate with objective measures of PA, with coefficients

ranging from r=0.3 to r=0.6. Based on this, a value of 0.5 was considered to be appropriate. Large measurement error increases variance in individual's scores, and this increased variance will reduce the strength of correlations between the IPAQ-SF and PASE.

The second hypothesis compared scores in PASE and IPAQ-SF within different domains of PA. Due to correlations within the same domains of PA having been conducted, it was expected that there would be correlations of at least r=0.5. Data were not normally distributed, Spearman's rank coefficient was used. Again, similar to the first hypothesis, r=0.5 was selected due to the expected measurement error.

The third construct validity hypothesis evaluated the discriminant construct validity of the IPAQ-SF and PASE. Chapter Seven's results found those reporting higher physical functioning also reported higher levels of PA. Previous studies have also shown higher levels of PA in those with increased self-reported physical health scores (Warburton et al., 2006; Powell et al., 2011; Yorston et al., 2012). The cut-off value of r=0.2 was used to reflect the correlation between physical functioning and PA levels in a previous study (Yorston et al., 2012). This is due to lower physical functioning found to be associated with lower levels of PA (Yorston et al., 2012), therefore it was anticipated that correlations between physical functioning and the IPAQ-SF and PASE total score would be at r=0.2 or greater within this sample.

The fourth construct validity hypothesis is that those who have fewer limitations in moderate activity are more likely to be physically active and it was felt that this should be reflected in the PASE and IPAQ-SF measurement of levels of PA

(Yorston et al., 2012). The fourth construct validity hypothesis was analysed using a one-way ANOVA with a Scheffe post hoc to identify differences in PASE total score for the different responders to the single item within the SF-12. In the IPAQ-SF, the residuals of the differences between groups were not normally distributed, and there was not homogeneity of variance. A one-way Kruskal-Wallis test was conducted with three independent samples and the Wilcoxon rank test was carried out as a post hoc analysis between the three groups, with alpha recalculated to $\alpha=0.05/3$, to control for type I error inflation using the Bonferroni correction.

9.4.5 Objective 3d

To assess and compare the responsiveness of total score and sub-domains in the IPAQ-SF and PASE.

Responsiveness analysis was conducted in responders who were classified were likely to have increased in PA compared to those who have not. Responders were classified as likely to have increased PA if:

- In the baseline questionnaire, responders reported not trying muscle strengthening exercises and general fitness exercises in the last three months.

  AND

- In the three month follow up questionnaire, responders reported in the past three months trying muscle strengthening exercises or general fitness exercises.

This meant that responders were deemed not to be taking part in regular PA at baseline but reported a change in their participation of PA at three months. Assessing responsiveness of the PASE and IPAQ-SF was conducted using data from the baseline and three month consultation questionnaires.

The responsiveness subsample change in total scores of the PASE and IPAQ-SF were calculated from baseline score to the three month follow up questionnaire. Responsiveness was evaluated using hypothesis testing. Box 9.2 summarises the hypothesis that was tested to evaluate responsiveness for objective 3d.

Box 9.2 Hypothesis tested for responsiveness in objective 3d

Hypothesis tested in objective 3d
The change in total score of the PASE will be positively correlated with the change in total IPAQ-SF score at a cut-off of r=0.5 or greater.

Correlations between IPAQ-SF and PASE change in total scores were conducted using Spearman's rank coefficient to evaluate responsiveness. Cut off values for correlations were *r*=0.5, which was taken to show agreement between IPAQ-SF and PASE for responsiveness. Responsiveness was also evaluated the subdomains of the PASE and IPAQ-SF: sitting activities, walking, moderate and vigorous intensity activities. To describe the changes in the responsiveness subsample over the three month period changes in BMI, joint pain intensity, SF-12 scores, EQ-5D scores and global assessment of change to joint pain were also reported.

Responsiveness was also evaluated in the IPAQ-SF and PASE using effect size (ES), standardised responsiveness mean (SRM), and responsiveness ratio (RR) (Paul et al., 2004). ES, SRM and RR were calculated for PASE and IPAQ-SF scores in the responsiveness subsample, and compared for comparisons of

responsiveness of both measurements. Comparisons of ES, SRM and RR have previously been used as measures of responsiveness in other studies (Paul et al., 2004; Packham et al., 2011).

ES was defined as the difference in mean change in the responsiveness subsample divided by the baseline standard deviation of the responsiveness sample (Kazis et al., 1989). The size of responsiveness for ES was interpreted using cut-offs described by Cohen (1977): small (at least 0.2); medium (at least 0.5); and large (at least 0.8 or greater).

SRM was calculated as the mean change in the subsample divided by the standard deviation of change scores in the subsample (Liang et al., 1990). SRM values were interpreted using the same cut-off values of Cohen (1977) and Liang (Liang et al., 1990).

RR was calculated as the mean difference between baseline and 3-month follow up in the responsiveness subsample, divided by the baseline standard deviation of the reliability subsample (Guyatt et al., 1987). A RR value > 1 indicates the ability of an instrument to detect change above a group with stable levels of PA (Guyatt et al., 1987).

9.5 Results

9.5.1 Results of data cleaning

When examining the IPAQ-SF total scores, no issues of outliers were identified. In the PASE, two outliers were identified within the three month dataset and these

two cases were examined. In one of those cases the scores were adjusted to reflect likely level of PA as the responder had reported the same activity multiple times. This was modified to be included only once, changing that responders total PASE score from 509 to 235. In the other case with a reported PASE total score of 480, indicating a high amount of PA, the respondent had given examples that suggested the score was a true reporting of their PA and so remained the same.

9.5.2 Characteristics of the population

In the baseline consultation questionnaire, 641 participants were mailed out a survey and 525 (81.9%) responded. Of the 525 who responded to the baseline consultation questionnaire, 432 (82.3%) completed the PASE questionnaire and 367 (69.9%) completed the IPAQ-SF. At baseline, 310 (59.0%) completed both the IPAQ-SF and PASE of the consultation questionnaire.

There were 524 3-month follow up consultation questionnaires mailed and 455 (86.8%) responded. Of these, 379 (83.3%) participants completed the PASE questionnaire and 339 (74.5%) completed the IPAQ-SF. In the three month follow up consultation questionnaire, 295 (64.8%) completed both the PASE and the IPAQ-SF. Of the 455 responders to the baseline and three month follow up consultation questionnaire, 324 (71.2%) completed the PASE questionnaire at baseline and at three months, with 264 (58.0%) who completed the IPAQ-SF at baseline and three months. In this study baseline data was taken from 432 respondents of the PASE and 367 respondents of the IPAQ-SF. Three month data was taken from the 379 respondents of the PASE and the 339 respondents of the IPAQ-SF. Figure 9.1 displays a flowchart of the response rate to the MOSAICS baseline and three month consultation questionnaire. Table 9.2 below compares

the baseline characteristics of those that had completed either the IPAQ-SF or PASE at baseline and the responders to the baseline consultation questionnaire but who did not complete either the IPAQ-SF or PASE at baseline.

Figure 9.1 Flowchart of MOSAICS baseline and three months survey

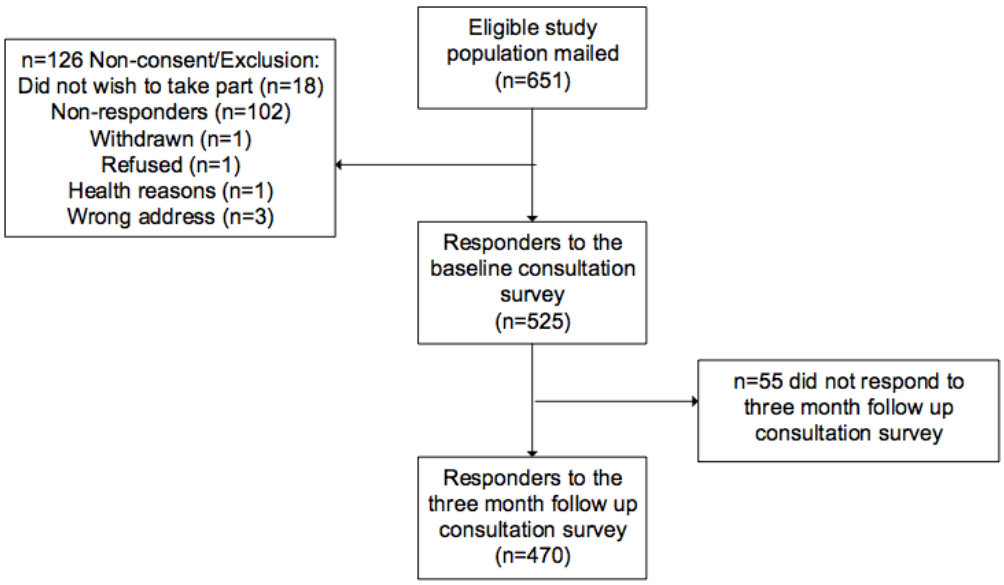


Table 9.2 Baseline characteristics of respondents and respondents that did not complete the IPAQ-SF or PASE.

	Respondents (n=489)	Respondents did not complete IPAQ or PASE (n=36)	p value
Gender, n (%): Males Females	 199 (41.1) 285 (58.9)	 11 (32.4) 23 (67.6)	 0.310
Median age, years (interquartile ranges):	67.00 (60-75)	74 (66.5-79)	<0.001*
Age range, years:	46-92	47-86	
Median IMD deprivation score (interquartile ranges):	20354 (15526-31622)	19845 (15994-28273)	0.996
IMD deprivation score range:	522-32468	1265-32468	
Median BMI, kg/m^2 (interquartile ranges):	27.57 (24.90-31.20)	26.49 (23.38-31.62)	0.770
Median highest pain intensity in last month (0-10) (interquartile):	8 (6-9)	8 (6-9)	0.730
Median SF-12 health status (interquartile ranges): Physical component score Mental component score	 36.45 (28.41-45.10) 53.23 (43.79-59.42)	 35.63 (37.61-44.53) 48.08 (41.10-54.24)	 0.945 0.024*
Median EQ-5D score (interquartile):	0.69 (0.59-0.73)	0.69 (0.2-0.8)	0.962

Key: IMD = index of multiple deprivation, BMI = body mass index, SF-12 = short form 12 items, EuroQoL 5 dimensions, *statistically significant (p=<0.05) using independent sample Wilcoxon test.

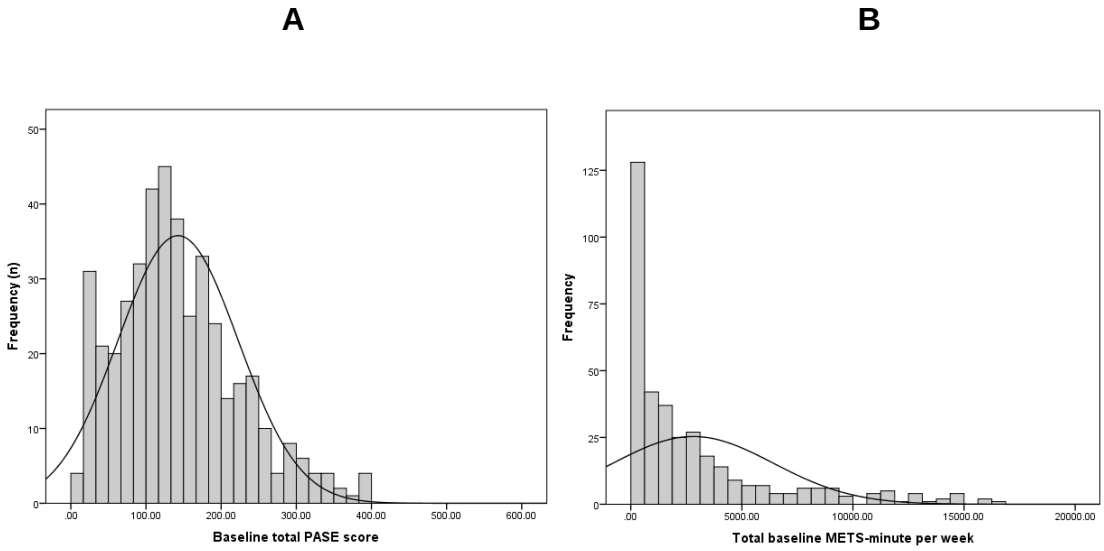
Those who completed the IPAQ-SF or PASE at baseline were significantly younger in age and had significantly better mental health scores in SF-12 compared to responders to the consultation questionnaire who did not complete either the IPAQ-SF or PASE. No significant differences were found in any other baseline characteristics.

9.5.3 Objective 3a

To compare the responses of individual items and total scores between the IPAQ-SF and PASE.

Figure 9.2 (A) displays the total score data of the PASE at baseline were normally distributed. The baseline PASE total score were not precisely normally distributed, although the distribution curves were symmetrically shaped. The baseline continuous total score of the IPAQ-SF was not normally distributed. Figure 9.2 (B) displays the baseline IPAQ-SF total score was positively skewed with n=125 (34.6%) of responders reporting zero $METS^{-1min-1week}$ at three month follow up. Appendix 1.11 and 1.12 display histograms of baseline and three months score in the PASE and IPAQ-SF.

Figure 9.2 Histogram of baseline total PASE scores (A) and IPAQ-SF total scores (B) and line of distribution



9.5.4 Objective 3b

To assess and compare the reliability and measurement error of the IPAQ-SF and the PASE.

This section describes the results of the reliability and measurement error analysis of the IPAQ-SF and PASE. Reliability and measurement error was assessed in a subsample (n=290) deemed to be stable in their level of PA at baseline to the three month follow up. Descriptive statistics are displayed in table 9.3 for the reliability and measurement error sub-sample at baseline and three month follow up. This information was used to estimate if the sub-sample was likely to have remained stable in levels of PA during the period from baseline to three month follow up.

Table 9.3 Descriptive statistics at baseline and three month follow up of reliability sub-sample

	Baseline	3 **month follow up**	**p value**
Gender, n (%): Males Females Missing	 114 (39.3) 174 (60.0) 2 (0.7)		
Mean age, years (sd):	68.34 (10.60)		
Age range, years:	46-92		
Median IMD deprivation score (interquartile):	20354 (15805.3-28099)		
IMD deprivation score range:	1040 - 32468		
Mean BMI, kg/m^2 (sd):	28.35 (4.86)	28.21 (4.90)	p=0.19
Mean highest pain intensity in last month (0-10) (sd):	7.42 (1.97)	6.14 (2.50)	p=<0.001*
Mean health status (sd): Physical component score Mental component score	 36.77 (11.51) 51.22 (10.87)	 37.81 (11.99) 51.35 (10.64)	 p=0.033* p=0.810
Mean EQ-5D score (sd):	0.58 (0.28)	0.63 (0.27)	p=<0.001*
Global assessment of change to joint problem(s), n (%): Missing Completely recovered Much better Better No change Worse Much worse		 3 (1.0) 26 (9.0) 54 (18.6) 109 (37.6) 77 (26.6) 11 (3.8) 10 (3.4)	
Mean PASE score (sd):	140.27 (76.34)	131.50 (81.31)	p=0.043*
Median IPAQ-SF total $METS^{-1min-1week}$ (interquartile range):	1386 (198-3451.5)	1708.50 (246-3564)	p=0.522+
IPAQ-SF categories, n (%): Low Moderate High	 64 (37.9) 56 (33.1) 49 (29.0)	 68 (40.3) 46 (27.2) 55 (32.5)	p=0.394‡‡

Key:*statistically significant below (p=<0.05) using paired *t*-test analysis. +not statistically significant using paired sample Wilcoxon signed rank test. ‡‡ not statistically significant below (p=<0.05) using McNemar test. sd equals standard deviation. Categorical variables percentages are displayed running down columns.

The reliability sub-sample did not change statistically in mean BMI, mean mental component score of SF-12 and median IPAQ-SF total METS$^{-1min-1week}$ between baseline and three month follow up, as shown in table 9.3. The sub-sample did change significantly, in terms of mean highest joint pain intensity, PCS and EQ-5D score. Total PASE score between baseline and three month follow up decreased significantly, showing the only decrease within the reliability sub-sample. The highest frequency of responders in this subsample reported at three months that their joint problem(s) had not changed in the global assessment of change. Only 29 (10%) had reported getting much better or completely recovering and 10 (3.4%) reported getting worse. The statistically significant changes in the subsample for joint pain intensity, health status and EQ-5D were not clinically meaningful changes (Jenkinson et al., 1997; Ware. 2000; Salaffi et al., 2004; Walters and Brazier, 2005).

In the reliability sub-sample the median IPAQ-SF score did not change significantly from baseline to three month follow up. The PASE had changed statistically significantly, with the PASE score decreasing at three month compared to baseline, suggesting a decrease in PA levels. The mean change in PASE scores was 8.74±62.0 from baseline to three months follow up.

9.5.5 Objective 3b - Reliability and measurement error of PASE total score

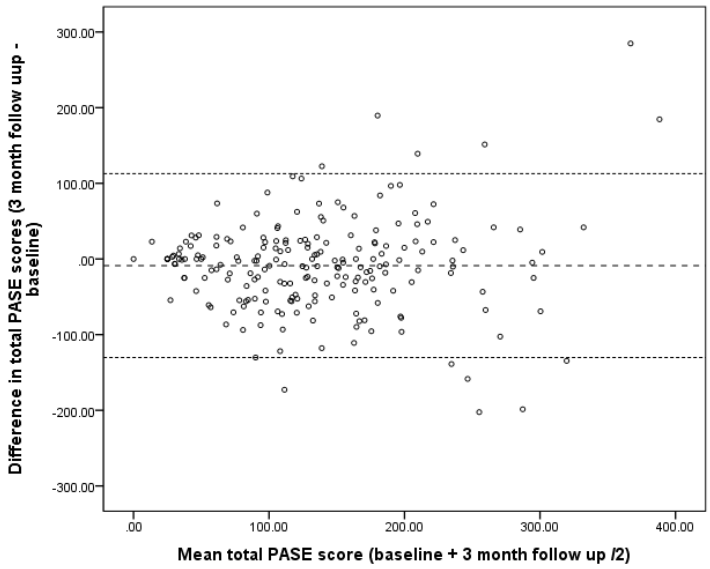
Table 9.4 displays the descriptive statistics of the reliability sub-sample that completed the PASE at baseline and three month follow up. Figure 9.3 below displays a Bland and Altman plot for the total score of the PASE at baseline and three months and is accompanied by the ICC statistical test for the PASE's total score.

Table 9.4 Descriptive statistics at baseline and three month follow up of reliability sub-sample that completed the PASE

	Baseline	3 Month follow up	p value
Gender, n (%): Males Females Total Missing	 83 124 207 1		
Mean age, years (sd):	67.48 (10.69)		
Age range, years:	46-92		
Median IMD deprivation score (interquartile ranges):	20880.00 (9353-23407)		
IMD deprivation score range:	1378-32468		
Mean BMI, kg/m^2 (sd):	28.43 (4.73)	28.23 (4.81)	0.10
Mean highest pain intensity in last month (0-10) (sd):	7.42 (1.98)	6.09 (2.49)	<0.001*
Mean health status (sd): Physical component score Mental component score	 36.98 (11.57) 52.22 (10.50)	 38.37 (11.76) 51.86 (10.66)	 0.01* 0.56
Mean EQ-5D score (sd):	0.60 (0.27)	0.64 (0.28)	0.01*
Global assessment of change to joint problem(s), n (%): Missing Completely recovered Much better Better No change Worse Much worse		 4 (1.9) 3 (1.4) 22 (10.6) 38 (18.3) 77 (37.0) 57 (27.4) 7 (3.4)	
Mean PASE score (sd):	140 (76.74)	131.50 (81.31)	0.043*

Key:*statistically significant below (p=<0.05) using paired *t*-test analysis. sd equals standard deviation. Categorical variables percentages are displayed running down columns.

Figure 9.3 Bland and Altman scatterplot of baseline total PASE scores and three month PASE total scores in reliability sub-sample



Key: N=208, darker circles on scatterplot signify multiple responders at same point. Middle dotted line = -8.76, displays mean difference between baseline and 3 month follow up PASE total score. Smaller dotted line display 95% limit of agreement, lower line = -130.28 and upper line = 112.76.

Table 9.4 displays that of the 290 respondents in the reliability sub-sample, 82 (28.3%) did not complete the PASE at baseline and three month follow up and were excluded from the reliability and measurement error analysis of the PASE. In the reliability sub-sample (n=208) that had completed the PASE at baseline and three months follow up, a two way random absolute agreement with ICC was conducted to assess reliability. The PASE did not quite reach adequate reliability in the test retest assessment (ICC=0.69, 95%CI= 0.61-0.76, p=<0.001) although the ICC for the PASE total score was significant with ICC. In evaluation of measurement error, SEM agreement for the PASE's total score was 46.74. The SDC for the total PASE was calculated to ±129.57. Figure 9.3 show the upper limit of agreement as 112.76 and lower limits of agreement as -130.28.

9.5.6 Objective 3b - Reliability of PASE subdomains

The data in each of the leisure time activities subdomains was not normally distributed and so a Spearman's rank coefficient was conducted to estimate reliability. Table 9.5 displays the Spearman's rank coefficient for each of leisure time activities in the PASE. Table 9.6 displays the limits of agreement in items within the household activities and the work-related activities domains.

Table 9.5 Reliability of PASE leisure time activities subdomains

Domain	Spearman's rank coefficient (r=)	Spearman's Rank coefficient p value
Sitting activities	0.54	<0.001
Walking activities	0.57	<0.001
Light intensity activities	0.33	<0.001
Moderate intensity activities	0.47	<0.001
Strenuous intensity activities	0.64	<0.001
Muscle endurance or strengthening activities	0.56	<0.001

In those who reported having worked or volunteered in work in the past seven days at baseline and three months (n=41), a two-way random absolute agreement ICC for the number of hours spent working or volunteered work in the past seven days was ICC=0.86, 95%CI= 0.75-0.92 (p=<0.001), suggesting this item showed an adequate level of reliability. Table 9.6 shows that agreement between baseline and three month follow up in items on household activities and work-related activities varied, ranging from *K*=0.37-0.65, this represents a fair to moderate level of agreement in these items (Landis & Koch, 1997).

Table 9.6 Kappa agreement in household activities, and the work-related activities domains of the PASE

Household activities	Kappa (*K*)	Kappa 95% CI	Interpretation of Kappa
Have done light housework in past 7 days, n (%):	0.49	0.08-0.89	Moderate
Have done heavy housework in past 7 days, n (%)	0.60	0.48-0.72	Substantial
Engaged in home repairs in past 7 days, n (%)	0.49	0.32-0.65	Moderate
Engaged in lawn work or garden care in past 7 days, n (%)	0.46	0.34-0.58	Moderate
Engaged in outdoor gardening in past 7 days, n (%)	0.37	0.25-0.50	Fair
Caring for another person in past 7 days, n (%)	0.52	0.39-0.64	Moderate
Work-related activities			
Worked or volunteered work in past 7 days, n (%)	0.65	0.53-0.77	Substantial
In those that worked in past 7 days, type of physical activity best describe work:	0.50	0.29-0.71	Moderate

95% confidence intervals (95% CI) calculated using Fleiss & Cohen (1973) method using http://faculty.vassar.edu/lowry/kappa.html. Interpretation of Kappa values using Landis & Koch (1997).

9.5.7 Objective 3b - Reliability and Measurement error of IPAQ-SF continuous total score

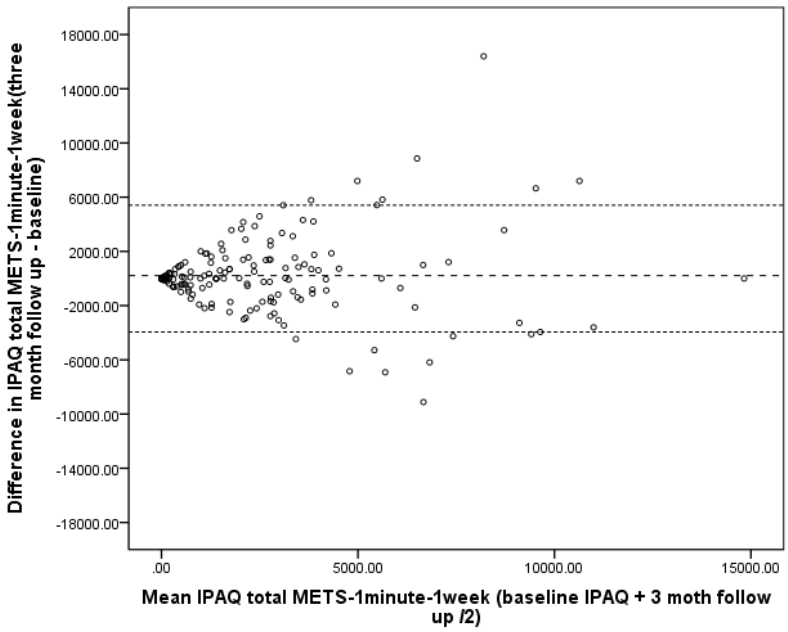
The IPAQ-SF total score, given in $METS^{-1minute-1week}$ data were not normally distributed at both baseline and three month follow up. Reliability was assessed for the IPAQ-SF total $METS^{-1minute-1week}$ using non-parametric methods. Table 9.7 displays the descriptive statistics of the reliability subsample that completed the IPAQ-SF at baseline and three month follow up.

Table 9.7 Descriptive statistics at baseline and three month follow up of reliability sub-sample that completed the IPAQ-SF

	Baseline	3 Month follow up	p value
Gender, n (%): Males Females Total	 71 (42.00) 98 (58.00) 169 (100)		
Mean age, years (sd):	67.95 (10.70)		
Age range, years:	47-92		
Median IMD deprivation score (interquartile ranges):	20950 (9878-32022)		
IMD deprivation score range:	2078- 32468		
Mean BMI, kg/m^2 (sd):	28.11 (4.65)	27.99 (4.59)	0.154
Mean highest pain intensity in last month (0-10) (sd):	7.54 (1.91)	6.19 (2.53)	<0.001
Mean health status (sd): Physical component score Mental component score	 36.48 (11.91) 52.17 (10.79	 37.76 (12.08) 51.92 (10.97)	 <0.001 <0.001
Mean EQ-5D score (sd):	0.59 (0.27)	0.62 (0.29)	<0.001
Global assessment of change to joint problem(s), n (%): Missing Completely recovered Much better Better No change Worse Much worse		 4 (2.4) 2 (1.2) 15 (8.9) 32 (18.9) 62 (36.7) 46 (27.2) 8 (4.7)	
Median IPAQ-SF total $METS^{-1min-1week}$ (interquartile range):	1386.00 (0.00-4639)	1708.50 (0.00-5080.50)	0.522[+]
IPAQ-SF categories, n (%): Low Moderate High	 64 (37.9) 56 (33.1) 49 (29.0)	 68 (40.2) 46 (27.2) 55 (32.6)	 0.394[++]

Key:[*]statistically significant below (p=<0.05) using paired *t*-test analysis. [+]not statistically significant using paired sample Wilcoxon signed rank test. [++] not statistically significant below (p=<0.05) using McNemar test. sd equals standard deviation. Categorical variables percentages are displayed running down columns.

Figure 9.4 Bland and Altman scatterplot of baseline and three month follow up total METS$^{-1minute-1week}$ scores from IPAQ-SF



Darker circles on scatterplot signify multiple responders at same point. Middle dotted line = -214.87, displays mean difference between baseline and 3 month follow up PASE total score. Smaller dotted line displays 95% ranked limit of agreement, Lower line = -3942 METS$^{-1minute-1week}$, Upper line = 5409 METS$^{-1minute-1week}$.

Of those selected for the reliability subsample, 122 (42.07%) did not complete the IPAQ-SF at baseline and three month follow up and were excluded from the reliability and measurement error analysis leaving 168 respondents. In the reliability sub-sample that had completed the IPAQ-SF at baseline and three months follow up, the Spearman's rank coefficient was r=0.58 (p=<0.01). Figure 9.4 displays the agreement using non-parametric method for limits of agreement (Bland & Altman, 1998). The upper ranked 95% limit of agreement for the IPAQ-SF in the reliability sub-sample was 4509METS$^{-1minute-1week}$ and the lower ranked 95% limit of agreement was -3942METS$^{-1minute-1week}$. This suggests that scores in IPAQ-SF need to change or be different by +4509METS$^{-1minute-1week}$, or -3942METS$^{-1minute-1week}$.

9.5.8 Objective 3b - Reliability of IPAQ-SF subdomains

Four subdomains of the IPAQ-SF were evaluated for reliability: time spent participating in sitting activities, $METS^{-1minute-1week}$ expended in walking activities, $METS^{-1minute-1week}$ expended in moderate intensity activities and $METS^{-1minute-1week}$ expended in vigorous intensity activities. Table 9.8 displays the Spearman's rank coefficient for each of the domains in the IPAQ-SF.

Table 9.8 Reliability of IPAQ-SF subdomains

Domain	Spearman's rank coefficient (r=)	Spearman's Rank coefficient p value
Sitting activities	0.74	<0.001
Walking activities	0.59	<0.001
Moderate intensity activities	0.45	<0.001
Vigorous intensity activities	0.46	<0.001

9.5.9 Objective 3b - Reliability of categorical IPAQ-SF score

Reliability of IPAQ-SF categorical data for IPAQ-SF at baseline and three months was analysed using level of agreement Table 9.8 above shows the responders' levels in each category of the IPAQ-SF at baseline and three months. A quadratic weighted Kappa showed agreement between baseline and three months of *K*=0.56 (95%CI= 0.43-0.67) which shows moderate agreement (Landis & Koch, 1997).

9.5.10 Objective 3c - Construct validity

To assess and compare the construct validity of total score and sub-domains in the IPAQ-SF and PASE.

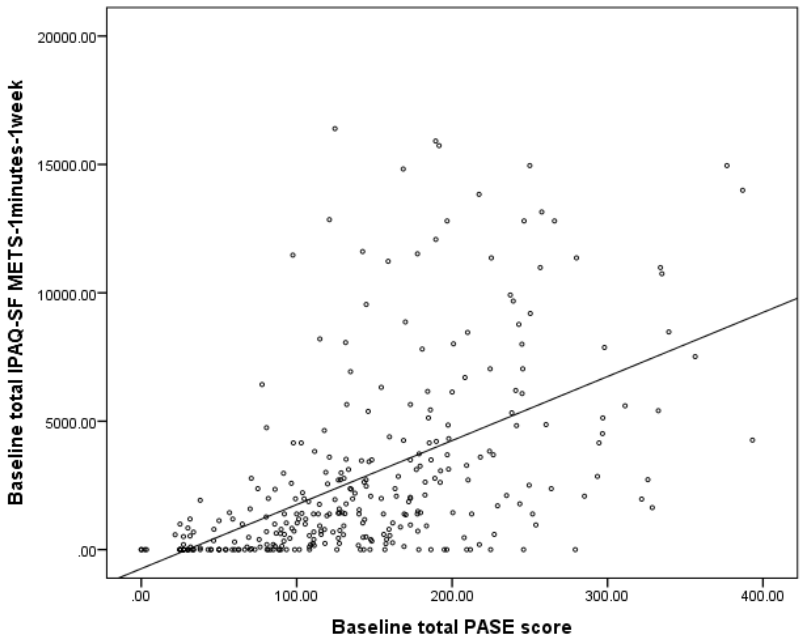
(1) Construct validity hypothesis one

The first hypothesis of construct validity was:

The baseline total score of the PASE will be positively correlated with the baseline total IPAQ-SF score at a cut-off of r=0.5 or greater.

Figure 9.5 displays a scatterplot of the baseline IPAQ-SF total METS$^{-1minute-1week}$ score and PASE total score.

Figure 9.5 Scatterplot of baseline total PASE scores and IPAQ-SF METS$^{-1minute-1week}$ scores



Key: Darker circles on scatterplot signify multiple responders at same point.

In responders who had completed the IPAQ-SF and PASE at baseline (n=310), there was a significant correlation using Spearman's rank correlation between the baseline PASE total score and IPAQ-SF total METS$^{-1min-1week}$, *r*=0.62, (two-tailed) p=<0.01. Figure 9.5 displays the range of variance in PASE and IPAQ-SF total score for individuals. Baseline mean PASE total score was 145.09±79.46, the median IPAQ-SF total METS$^{-1min-1week}$ was 14440.00 (206.25-3600)METS$^{-1min-1week}$ (interquartile ranges). Figure 9.5 shows a large number of respondents with a zero score for the IPAQ-SF while reporting a range of scores in the PASE.

(2) Construct validity hypothesis two

The second hypothesis of construct validity was:

The baseline scores for time spent sitting, walking, moderately active and vigorously active subdomains of the IPAQ-SF and PASE will be positively correlated at r=0.5 or greater.

This hypothesis assessed construct validity in the subdomains of the IPAQ-SF and PASE. As the IPAQ-SF consists of four subdomains and the PASE consists of six leisure activities domains, household activity domain and work related activity domain, not all of the PASE domains were used. Table 9.9 displays the Spearman's rank coefficient of the baseline sitting activities, walking activities, moderate intensity activities and vigorous intensity activities of the IPAQ-SF and PASE.

Table 9.9 Correlations between subdomains of the IPAQ-SF and PASE

Domain	Spearman's rank coefficient (r=)	Spearman's Rank coefficient p value
Sitting activities	0.46	<0.001
Walking activities	0.57	<0.001
Moderate intensity activities	0.34	<0.001
Vigorous intensity activities	0.39	<0.001

Table 9.9 indicates that while walking activities correlate at greater than r=0.5 other sub domains of the IPAQ-SF and PASE did not correlate sufficiently for hypothesis two.

(3) Construct validity hypothesis three:

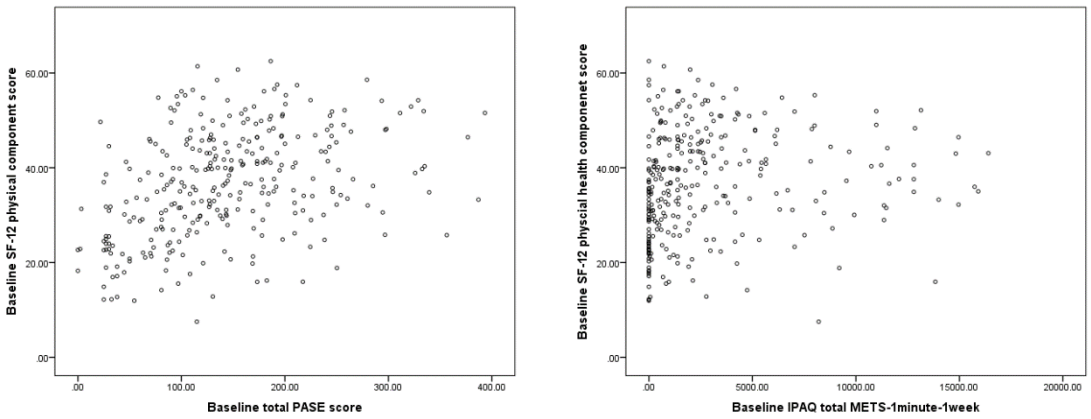
The third hypothesis of construct validity was:

The IPAQ-SF and PASE total scores will correlate positively with the physical health component of the SF-12.

This hypothesis was to evaluate the PASE and IPAQ-SF discriminant construct validity and ability to show relationships between other constructs and levels of PA. Figures 9.6 (A) and 9.6 (B) below show scatterplots of total scores of the PASE and IPAQ-SF against SF-12 physical health component scores.

Figure 9.6 Scatterplot of baseline PASE total score (A) and IPAQ-SF (B) and SF-12 physical health component scores

(A) **(B)**



Darker circles on scatterplot signify multiple responders at same point.

Of the responders who had completed the PASE and IPAQ-SF at baseline, seven (2.3%) did not complete the SF-12 so were excluded. In responders who had completed the SF-12 and PASE at baseline (n=303), there was a significant correlation using Pearson's correlation between baseline PASE total score and SF-12 PCS, r=0.39, (two-tailed) p<0.01. There was also a significant correlation using Spearman's rank correlation between baseline IPAQ-SF total METS-1minute-1week score and SF-12 PCS, r=0.30, (two-tailed) p<0.01. Both these correlation coeiffients were greater than the cut-off value set within the hypothesis.

(4) Construct validity hypothesis four:

The fourth hypothesis of construct validity was:

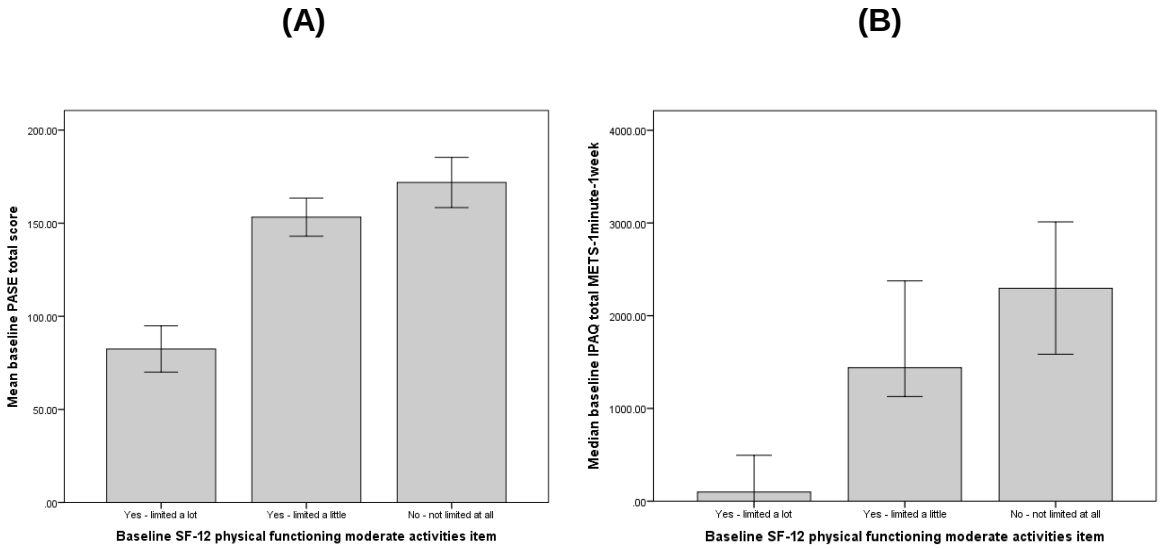
Respondents reporting more limitation in moderate activities in the SF-12 will have significantly lower levels of PA in the PASE and IPAQ-SF compared to those reporting less limitation.

Figure 9.7 (A) and 9.7 (B) display bar charts of IPAQ-SF total $METS^{-1minute-1week}$ and PASE total score in the three different responses to the SF-12 single item.

(1) PASE

In responders who completed the SF-12 and PASE at baseline (n=303), there was a significant difference in the PASE score in those who reported different levels of physical function, using a one way ANOVA model: F= 46.27; df= 2, 427; p=<0.001. As shown in figure 9.7 (A) the mean baseline PASE total score was statistically significantly lower in those with ‘a lot’ of limitation compared to those with ‘little’ limitation (p=<0.001) and those with ‘no’ limitation (p=<0.001) using a Scheffe post hoc analysis. There was no statistically significant difference between those with a ‘little’ limitation and those with ‘no’ limitation (p=0.73) in mean PASE total score with the Scheffe post hoc analysis. Figure 9.7 (A) shows a small difference in mean PASE total score between those with a ‘little’ limitation and those with ‘no’ limitation.

Figure 9.7 Bar chart of baseline PASE total activity score and SF-12 moderate activity single item



Key: Error bars represent 95% CI

(2) IPAQ-SF

In responders who had completed all of the SF-12 and IPAQ-SF at baseline (n=303), there was a significant difference in IPAQ-SF score in those who reported different levels of physical function, using Kruskal Wallis test: h= 88.91; *df*= 2; *p*=<0.001. As shown in figure 9.7 (B) the median baseline IPAQ-SF total score was statistically significantly lower in those with ‘a lot’ of limitation compared to those with ‘a little’ limitation (*p*=<0.001) and those with ‘no’ limitation (*p*=<0.001) using a Bonferroni-corrected Wilcoxon rank sum test (α= 0.05/3 = 0.167). There was not a statistically significant difference between those with ‘a little’ limitation and those with ‘no’ limitation (*p*=0.40) in median IPAQ-SF total score within the Bonferroni-corrected Wilcoxon rank sum test.

9.5.11 Objective 3d - Responsiveness

To assess and compare the responsiveness of total score and sub-domains in the IPAQ-SF and PASE.

In all responders there were 61 (13.4%) (n=455) that fitted the responsiveness sub-sample criteria and of those in the responsiveness sub-sample, 34 (55.7%) completed the IPAQ-SF at baseline and the three month follow up and 46 (75.4%) completed the PASE at baseline and three month follow up. There were 52 of the responsiveness sub-sample who had completed the IPAQ-SF, PASE or both at baseline and three month follow up. Table 9.10 displays descriptive statistics of the responsiveness sub-sample at baseline and three month follow up to indicate if the responsiveness sub-sample were likely to have changed their PA behaviour during the three month period.

Table 9.10 Baseline and three month follow up descriptive statistics of responsiveness sub-sample

	Baseline	3 Month follow up	p value
Gender, n (%): Males Females Missing	 27 (52.9) 24 (47.1) 1		
Mean age, years (sd):	66.76 (9.87)		
Age range, years:	46-86		
Median IMD deprivation score (interquartile range):	20182 (15989-26948)		
IMD deprivation score range:	723 – 32468		
Mean BMI, kg/m^2 (sd):	29.7 (6.21)	29.18 (6.28)	p=0.18
Mean highest pain intensity in last month (0-10) (sd):	6.67 (2.26)	6.13 (2.69)	p=0.06
Mean health status (sd): PCS MCS	 38.10 (9.19) 51.75 (11.28)	 37.72 (10.56) 51.15 (10.59	 p=0.70 p=0.60
Mean EQ-5D score (sd):	0.60 (0.28)	0.63 (0.26)	p=0.34
Global assessment of change to joint problem(s), n (%): Completely recovered Much better Better No change Worse Much worse Missing		 1 (1.9) 3 (5.8) 11 (21.2) 23 (44.2) 11 (21.2) 3 (5.8) 0	
Mean PASE score (sd):	153.42 (89.26)	139.05 (83.52)	p=0.17
Median IPAQ-SF total $METS^{-1min-1week}$ (interquartile range):	2574.0 (305.3-5152.9)	2227.0 (396-4986)	p=0.14
IPAQ-SF categories, n (%): Low Moderate High	 8 (23.5) 11 (32.4) 15 (44.1)	 10 (29.4) 9 (26.5) 15 (44.1)	 P=0.532

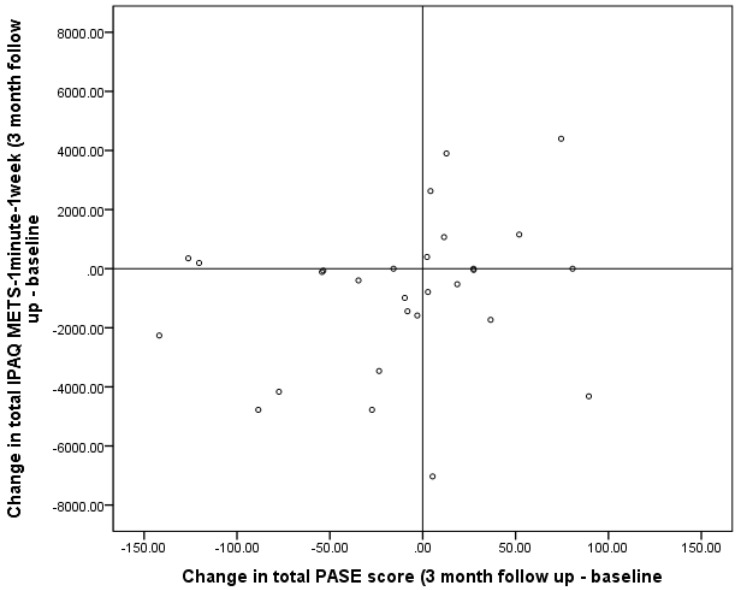
Key: sd equals standard deviation. In categorical variables for gender and IPAQ-SF category percentages are shown going down the columns. Mean significance tests using paired sample t tests. Median significance tests using Wilcoxon paired rank sum test. Categorical difference not significant using McNemar test.

Table 9.10 displays descriptive statistics from baseline and three month follow up in the responsiveness sub-sample. The change in mean PASE scores (-14.38) and change in mean IPAQ-SF score (-579.28METS$^{-1min-1week}$) were negative changes, suggesting a reduction in levels of PA, although changes were not statistically significantly different. The change in mean PASE scores was smaller in the responsiveness sub-sample compared to that of the SDC as shown in section 9.5.5, as well as being within the SEM and limits of agreement, suggesting this change in score may be due to measurement error. Changes in median IPAQ-SF score were also found to be within the 95% ranked lower limits of agreements suggesting no 'true' changes in PA outside of measurement error.

The ES for the PASE was -0.16, as change scores in the responsiveness subsample decreased in the PASE from baseline to 3 months. Similarly, the ES for the IPAQ-SF was -0.14. SRM for the PASE was -0.21 and for the IPAQ-SF was -0.21. The RR for the PASE was 0.09 and IPAQ-SF was 0.12.

A lack of changes in other variables suggests that the responsiveness sub-sample selected may not have changed in PA behaviour during the three months. No significant changes in pain intensity, SF-12 PCS and MCS or EQ-5D were found as shown in table 9.9. The highest percentage of the responsiveness sub-sample also reported that there was no change in their joint problem in the global assessment of change at three months, further suggesting this cohort was stable. Figure 9.8 displays a scatterplot of change in scores for IPAQ-SF and PASE in the responsiveness sub-sample.

Figure 9.8 Scatterplot of baseline to three month follow up change in IPAQ-SF total METS$^{-1minute-1week}$ score and change in total PASE score in responsiveness sub-sample



Key: Darker circles on scatterplot signify multiple responders at same point.

Figure 9.8 demonstrates the large amount of variance in the responsiveness subsample (n=34) in changes of levels in PA as measured by the PASE and IPAQ-SF in the three month period. The mean change in the PASE total score of the responsiveness subsample, shown in table 9.10, did not exceed SDC found in section 9.5.5. The responsiveness subsample's median IPAQ-SF change in METS$^{-1min-1week}$ did not exceed the upper limits of agreement shown in Section 9.5.7. This suggests that the changes in levels of PA in the responsiveness subsample would seem to be due to measurement error rather than a true change in levels of PA. It is also possible that the responsiveness subsample did not change its PA behaviour, or both the IPAQ-SF and PASE are not responsive to changes in this population's change in PA behaviour. A responsiveness analysis

of this sub-sample was not possible as indicators suggest this sample had not changed in PA behaviour.

9.6 Discussion

This section discusses the main findings in line with the overall aim of this study to evaluate the measurement properties of the International PA Questionnaire Short Form (IPAQ-SF) and the PA Scale for the Elderly (PASE) in the target population and in the context of each of the objectives as outlined in section 9.3. Comparisons of the findings in relation to other relevant research are made together with the strengths and limitations of this study; the recommendations for future research are not discussed in this section but in Chapter ten of the thesis.

9.6.1 Main Findings

Prior to this analysis of measurement properties of the IPAQ-SF and PASE, the systematic review in Chapter eight indicated that the IPAQ-SF and PASE were potentially the most appropriate measure of PA in the target population based on their measurement properties in adults aged 45 years and over with joint pain or OA and in older community dwelling adults. It was not clear what reliability, measurement error and validity of the IPAQ-SF and PASE would be in adults aged 45 years and over consulting primary care with joint pain. In addition, the systematic review in Chapter eight showed that no study had evaluated the responsiveness of the IPAQ-SF or PASE.

Using data taken from the baseline and three months MOSAICS consultation questionnaire, description of the IPAQ-SF and PASE distribution of total scores

and evaluation of IPAQ-SF and PASE reliability, measurement error, construct validity and responsiveness were conducted.

The main findings of this study were that the PASE data was normally distributed where as the IPAQ-SF were positively skewed, with 34.6% of respondents to the IPAQ-SF reporting $0METS^{-1min\text{-}1week}$; this represents a high number of respondents with a floor effect. It is not clear what the cause of this is in the IPAQ-SF but compared to the PASE a possible explanation could be that the IPAQ-SF does not include lower intensity activities of PA and household or work-related activities.

When evaluating the reliability and measurement error of the IPAQ-SF and PASE there were clear limitations to both instruments' ability to measure levels of PA in the target population reliably and without measurement error. While reliability for both the IPAQ-SF and PASE was below the cut-off value set, both instruments were not low in reliability. When evaluating measurement error, the IPAQ-SF and PASE both demonstrated large measurement error, this severely affects the ability of both instruments in measuring PA. For example, if an individual adult with joint pain had a true level of PA which consisted of participating in moderate intensity of PA for 30 minutes on 5 days a week, the weekly energy expenditure would be approximately $500METS^{-1min\text{-}1week}$. Due to the measurement error of the IPAQ-SF evaluated in this study, this individual's IPAQ-SF score could range from $0METS^{-1min\text{-}1week}$ to $5909METS^{-1min\text{-}1week}$ based on the 95% limits of agreement of the IPAQ-SF in figure 9.3. This situation in measurement error was found to be similar in the PASE, with measurement error of the PASE in limits of agreement and smallest detectable change both above a PASE score of 100. Given that most PASE

scores range from 0-400 (Washburn et al., 1999), this also represents a large error.

This large measurement error is important as it makes differentiating individual's level of PA difficult, it also makes describing a responder's level of PA using the IPAQ-SF or PASE difficult as mean IPAQ-SF or PASE scores will have large measurement error within them.

Given that there was found to be large measurement error it was not surprising that the evaluation of construct validity showed that the PASE and IPAQ only moderately correlate with each other, this suggests that both instruments are measuring the same construct of self-report PA. Both instruments showed the large measurement error, it also displays both measures are similarly limited to evaluating levels of PA in the target population. In discriminative validity, both the IPAQ-SF and PASE did display that the correlations with PA levels and physical health using the SF-12 were possible to be identified despite the large measurement error.

Evaluation of responsiveness was conducted evaluating the ES, SRM and RR of the IPAQ-SF and PASE. Because of the large varieity in change scores in the IPAQ-SF and PASE, ES, SRM and RR were negative and low, suggesting the two instruments are poor at detecting change. The subsample selected to evaluate responsiveness did not appear to change, overall in PA behaviour during the three month period using the IPAQ-SF and PASE. The limitations of evaluating the responsiveness of the IPAQ-SF and PASE could be linked to that of the large measurement error and any 'true' change in levels of PA could not be detected.

When considering responsiveness of the IPAQ-SF and PASE the large measurement error in both instruments suggests that having a change outside of the measurement error would be so large that it would likely represent a meaningful change in ‘true’ PA levels for their joint pain. Due to the large measurement error identified in objective 3b, the IPAQ-SF and PASE are likely to be poor in detecting changes in levels of PA in the target population.

9.6.2 Appropriateness and feasibility of PASE and IPAQ-SF

This section evaluates the appropriateness and feasibility of the IPAQ-SF and PASE for measuring PA in the target population. In terms of feasibility of the IPAQ-SF and PASE, the rates in which the two instruments were completed with total PA level scores differed. The IPAQ-SF only had PA level scores for n=264 (58.0%) of responders, with a total PA score at both baseline and three months, compared to n=324 (71.2%) for the PASE. The reason for this poor completion rate may not be entirely due to patients not completing the IPAQ-SF. It could be that the IPAQ-SF contains an ‘unsure’ option for responders, to indicate that they were unsure about how to answer the items in the IPAQ-SF. If an individual selects this option, they are not allocated a score in the IPAQ-SF.

There is no unsure option within the PASE. Because of this, it had a higher number of completed questionnaires. This could be due to wording, as the PASE asks individuals about different aspects of their PA, which may have stimulated a response, compared to the IPAQ-SF wording of items, which did not.

A positive aspect of the PASE relates to how the items are structured within the instrument. Separating PASE responses into more independent dimensions of PA provides a description of how the PA of an individual is distributed. This allows for

exploring individuals' PA behaviour in a more detailed description of types of activities. This can be used to identify where changes in the distributions of different physical activities may occur, and where this does not affect the total score of the PASE. Appendix 1.13 illustrates the individual item responses of the PASE at baseline and 3 months. Separating out the PASE into individual items may show a more detailed summary of PA behaviours, compared to the total score alone.

9.6.3 Comparison with previous studies

The findings of this chapter are similar to those of previously published works on the measurement properties of the IPAQ-SF and PASE. Studies that have investigated the measurement properties of the PASE varied in methods but have similar findings (Svege et al., 2012; Bolszak et al., 2014). The results from this study found an ICC reliability of the PASE to be ICC=0.69, the reliability of the PASE was lower compared to other studies which had evaluated reliability of the PASE in OA populations. One study in Norway on hip OA adults reported nine day test-retest reliability in a sample size of 33 as ICC=0.78 (Svege et al., 2012). While another in Switzerland evaluated seven day test-retest reliability of the PASE in 25 males with a total knee arthroplasty reported an ICC of 0.77 and 25 females with a total knee arthroplasty an ICC of 0.58 (Bolszak et al., 2014). The variation in the ICC values may be explained by the fact that both of those studies had a smaller sample size compared to the reliability subsample in this one (33 in Svege et al. (2012) and 25 for males and females in Bolszak et al. (2014)). Both had shorter intervals between separate measurements compared to this study, in the work in Norway, the mean interval was nine days (Svege et al., 2012) and the study by

Bolszak et al. (2014) was seven days. In both studies, the authors had asked the participant to remain stable in their PA behaviour and it was assumed that in their sample the levels of PA remained stable.

Svege et al. (2012) found smaller measurement error and limits of agreement compared to the measurement error found in this study, but in line, reported that the measurement error in the PASE was large. Bolszak et al. (2014) also reported a large measurement error but reported it in percentages, which make it difficult to compare with Svege et al.'s (2012) study and this thesis.

The subdomains of the PASE were also evaluated for reliability (Svege et al., 2012) although the statistical methods used and scoring of the domains differed with this study. This study analysed individual items for household and work activities where the Svege et al.'s (2012) study examined total scores. This, and the hip OA population in the Svege et al.'s (2012) study, showed that reliability of the PASE's individual subdomains was lower compared to the PASE's total score. It is not clear why the reliability of the PASE's subdomains are lower compared to its total score. A rationale for this could be the target population recall the same total PA in the PASE, but allocate them differently in different items of the questionnaire between repeated measurements.

For the IPAQ-SF, an ICC was not possible to calculate as the data was not normally distributed for change scores. Reliability in the IPAQ-SF was found to be lower than the cut off value of r=0.7. Two studies on the reliability of the IPAQ-SF in sample of patients with total hip or knee arthroplasty have been conducted (Naal et al., 2009; Blikman et al., 2013) and demonstrated reliability varied. One study found a test retest range of r=0.49-0.81, in a nine to twelve day retest

interval in a relatively smaller sample of 44 participants (Blikman et al., 2013). The second study found much higher reliability in its sample and used ICC calculations to assess reliability (ICC=0.76-0.86), in a seven days retest interval and a sample of 79, which was larger compared to Blikman et al. (2013) and Naal et al's, 2009 studies. This indicates that the shorter intervals between test-retests in the work by Naal et al. (2009) may influence the higher reliability. In this study's reliability subsample of different sites of joint pain and three month follow up, reliability is low within the IPAQ-SF. Blikman et al. (2013) observed high measurement error and SDC in the IPAQ-SF, which was also apparent here with limits of agreement at -3942 and 5409 $METS^{-1minute-1week}$. To put this in context the amount of $METS^{-1minute-1week}$ of participating in moderate exercise 5 times a week for 30 minutes each time is equal to $500METS^{-1minute-1week}$ (Kaminsky, 2006; Bull & Expert Working Groups, 2010). This makes the limits of agreement very high in the analysis of the IPAQ-SF's measurement error. For an IPAQ-SF total score in this population to be separated without risk of measurement error, a change or difference in IPAQ $METS^{-1minute-1week}$ would need to exceed the large measurement error.

Only one previous study has compared the IPAQ-SF and the PASE in a construct validity assessment in the target population (Svege et al., 2012). The correlations of total scores were high between PASE and IPAQ-SF (Svege et al., 2012). Correlations between sub-domains of PA in the IPAQ-SF and PASE were lower in Svege et al., (2012), which support the findings here. Svege et al. (2012) also looked at the correlations of scores with the PASE to an objective measure of PA, using the Actigraph GT1M. Correlation to the objective measure and PASE was a Spearman rank correlation of r=0.30. This suggests that the PASE and IPAQ-SF are similar in their measurement of PA but both measures do not necessarily

reflect levels of PA in the target population, when compared with objectives measures.

9.7 Conclusion

From the findings of this study, and the systematic review in Chapter Eight, the use of self-report PA instruments, including IPAQ-SF and PASE, may not be sufficiently precise for the measurement of PA in adults with OA or joint pain, and not sensitive for detecting change. The IPAQ-SF or PASE would appear to have the same limitations as other self-report PA instruments identified in the systematic review of Chapter Eight. The IPAQ-SF and PASE were selected for this study as they both assess PA in terms of intensity, duration and frequency of PA, which is important for quantifying the amount of PA per week. When using the IPAQ-SF or PASE to measure PA in adults with joint pain or OA, attention should be given to the likelihood of large measurement error and the small correlation these self-report instruments have to more direct objective measurements of PA. The large measurement error in both instruments may be important, as it can increase the variance in levels of PA for samples where measurements are made. This may lead to Type I error when evaluating associations between PA and other measurements.

This chapter has examined the measurement properties of the IPAQ-SF and PASE in adults aged 45 years and over with joint pain or OA in the hip, knee, hand or foot. The main findings and conclusions of this chapter are that the IPAQ-SF and the PASE showed moderate reliability and construct validity but a large measurement error. The large measurement error identified in this study makes describing self-report levels of PA using either the IPAQ-SF or PASE limited in the

target population. Responsiveness of the IPAQ-SF and PASE could not be assessed in this study.

The next chapter will provide a comprehensive discussion of all of the findings from this thesis.

Chapter ten: Discussion

10.1 Overview

This chapter provides a summary of the rationale for the thesis, the three aims and the objectives. It describes the main findings together with an evaluation of the strengths and limitations of each study conducted to achieve these aims. The chapter also includes a discussion of the possible implications of the findings for future research and a final conclusion.

10.2 Physical activity and joint pain

PA has a number of physiological and psychological benefits, preventing many chronic health conditions and to prolong life (Blair et al., 1996). A physically active lifestyle is also important to maintain good health in the general population (Haskell et al., 2007; Bull & Expert Working Groups, 2010). A lifestyle of long periods of inactivity or a sedentary lifestyle are risk factors for many health conditions, biomarkers of poor health and factors that will decrease life expectancy (Rezende et al., 2014).

OA is the most common cause of musculoskeletal pain in adults aged 45 years and over (Lawrence et al., 2008), the most common sites being the knee, hip, hand and foot (Pereira et al., 2011). It is likely that adults aged 45 years and over with symptomatic joint pain in the knee, hip, hand or foot (the target population) either have already developed or are at risk of developing OA (Altman et al., 1986). Exercise reduces pain and disability in adults aged 45 years and over with OA; PA has been recommended by NICE, the Osteoarthritis Research Society

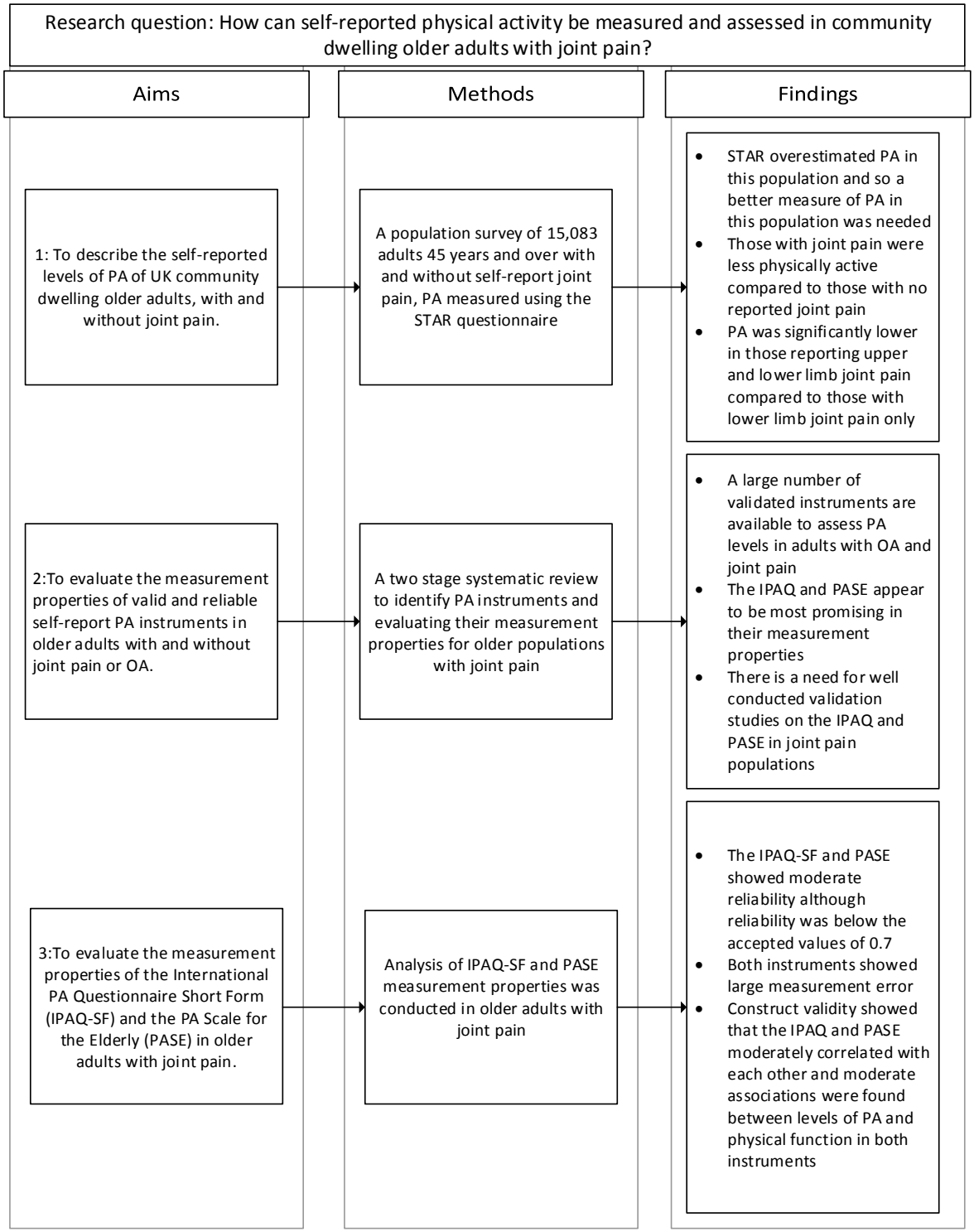
International (OARSI) and the American College of Rheumatology (ACR) in their most recent guidelines for the treatment of OA (Hochberg et al., 2012; McAlindon et al., 2014; NICE, 2014).

The number of adults aged 45 years and over who are physically active and those with low levels of PA, including sedentary life styles was previously unknown. Prior to undertaking this PhD, the most appropriate self-report measurement instrument of PA in adults aged 45 years and over with joint pain was also unclear.

10.3 Summary of key aims

Figure 10.1 illustrates each of the three aims addressed in this thesis. A short summary of the main findings in each study and how each objective links to the aims is also provided.

Figure 10.1 Summary flowchart of thesis aims, studies and main findings.



10.4 Main findings of thesis

The thesis contributed towards new knowledge on measuring and assessing self-reported PA in adults aged 45 years and over with joint pain. The following section details the main findings of the three aims of the thesis and compares findings with other related studies.

10.4.1 Aim one: To describe the self-reported levels of PA of UK community dwelling adults aged 45 years and over, with and without joint pain

The MOSAICS population survey demonstrated that adults aged 45 years and over with self-reported joint pain were less likely to be active compared to those with no self-reported joint pain. The STAR questionnaire (Matthews et al., 2005), appeared to overestimate PA levels in adults aged 45 years and over, both with and without joint pain in comparison to a larger UK adult population (BHF, 2012). In those with self-reported joint pain, the STAR classified 43.9% as active to UK Department of Health recommendations and in those with no joint pain, and 54.9% as active to recommendations. The BHF (2012) reported in similar aged general population adults, only 32% of males and 28% of females were active to the same recommendations. The MOSAICS population survey also found higher levels of PA in adults with joint pain compared to studies in OA populations from different countries (Dunlop et al., 2011a; De Groot et al., 2012; White et al., 2013).

The thesis identified that in the target population, physical and mental health was better in those with higher PA levels. The target population which reported low levels of activity in the STAR reported significantly lower physical and mental health compared with those reporting higher levels of PA. Those that reported low

levels of PA are a highly vulnerable group due to their lower PA levels and health status scores which is an indicator for risk to health conditions and premature death (Blair et al., 1996; Rezende et al., 2014). Those who stated being physically active to recommendations in the STAR (Bull & Expert Working Groups, 2010) had reported higher physical and mental health scores similar to those of general UK adult populations (Gandek et al., 1998). This demonstrates the association that PA can have as part of improving better health in adults aged 45 years and over with joint pain. Similar findings were found in a previous study which measured physical function in gait speed and levels of PA (Dunlop et al., 2011). It may be suggested that PA is associated with increased health and physical functioning in adults with joint pain, although other confounding factors alongside PA may play a role in health status and physical functioning, including age, BMI, socio-economic status and gender.

Adults aged 45 years and over with joint pain in the lower limbs only were compared to those with generalised joint pain (both upper and lower limbs) in terms of their PA levels. Levels of PA were lower in those with generalised joint pain compared to those with lower limb joint pain only. The implications of the findings within the generalised joint group mean this population is likely to report worse health outcomes. This finding is supported from other research that demonstrated that more sites of pain are associated with worse health outcomes (Kamaleri et al., 2008; Finney et al., 2013). This also represents an opportunity for targeting this population as one that could particularly benefit from increasing their PA levels.

10.4.2 Aim two: To evaluate the measurement properties of reproducible self-report PA instruments in adults aged 45 years and over with and without joint pain or OA

Instruments that were used previously in the target population were firstly identified and then evaluated for their measurement properties. This differed from approaches in previous systematic reviews which only investiaged measurement properties of insturments (Terwee et al., 2011). The systematic review undertaken in this thesis utilised a two-stage study design: initially (Stage A), the first stage included a systematic search to identify all self-report PA instruments previously used in relevant research was conducted, then the second stage (Stage B) included an evaluation of the measurement properties of the identified instruments.

The systematic review was a comprehensive approach to retrieve articles on self-report instruments used in the target population. The systematic review also drew on a hybrid of systematic review methods already used to evaluate PROMS (Smith et al., 2005; Terwee et al., 2011). The approach of the systematic review adopted a two-stage strategy and took the COSMIN criteria when data was synthesised and appraisal of study quality (Terwee et al., 2010; Mokkink et al., 2011). The advantage of this was to use the two-stage strategy to retrieve a maximal number of instruments to evaluate together with scientifically developed criteria of assessing measurement properties that has been recommended by international initiative Core Outcome Measures in Effectiveness Trials (COMET) for selecting core outcomes (Williamson et al., 2012).

Searches were made in Embase, MEDLINE and Web of Science databases. Data on each instrument's measurement properties were extracted using the QAPAQ checklist. Study methodology was assessed for quality using a modified COSMIN checklist. At least two reviewers independently appraised selection criteria, completed data extraction and methodology quality assessments.

The first stage identified 23 reproducible self-report instruments that had been previously used with adults with OA or joint pain. In the second stage, an evaluation of the measurement properties of identified instruments for the target population and community dwelling adults aged 45 years and over clarified that the IPAQ and PASE were the promising measures of PA; both evaluated PA in terms of frequency, duration and intensity of physical activities. They also displayed evidence of reliability in OA populations (table 8.32), and in studies of older community dwelling adults, the IPAQ and PASE demonstrated positive scores in reliability, construct and criterion validities (table 8.33). The systematic review concluded that the IPAQ and PASE appear promising instruments based on their measurement properties. There was not enough evidence currently recommending the IPAQ or PASE as valid for use in the target population and further investigation of their measurement properties was required; critically, responsiveness had not been assessed in any of the self-report instruments. Further investigation of the measurement properties of the PASE and IPAQ was important to undertake as the first stage found a large number of studies in joint pain or OA research used the IPAQ or PASE as an outcome measure of self-reported levels of PA.

10.4.3 Aim three: To evaluate the measurement properties of the International PA Questionnaire Short Form (IPAQ-SF) and the PA Scale for the Elderly (PASE) in adults aged 45 years and over with joint pain

The analysis of measurement properties evaluated the measurement properties of the IPAQ-SF and PASE in adults aged 45 years and over with joint pain using the MOSAICS consultation questionnaire at baseline and three month follow up.

At baseline completion responders to the consultation questionnaires were: n=339 (74.5%) for the IPAQ-SF and n=379 (83.3%) for the PASE. The IPAQ-SF contained an option that allowed responders to indicate if they were unsure or did not know their level of PA and this decreased completion rates for the instrument. For some who reported not knowing their level of PA on the IPAQ-SF, the PASE had been completed. Of those who had a total score for the IPAQ-SF at baseline, 34% reported 0 total $METS^{-1minute-1week}$ this represents a large floor effect and as a result of this IPAQ-SF scores were positively skewed. Problems in the floor effect of the IPAQ-SF are that the variance in those individuals is limited and can threaten the validity of research due to the zero variation at the low end of the IPAQ-SF scoring range. Floor effect can also affect research on interpreting interaction effects as the failure to find significant differences between the IPAQ-SF and another independent variable (De vet et al., 2011).

The IPAQ-SF and PASE both had moderate reliability, although the total score of PASE at baseline and three month follow up was below an intraclass correlation cut-off value (0.7) which signifies an acceptable level of reliability. The IPAQ-SF and PASE both demonstrated large measurement error in their measurement of PA. The upper 95% limit of agreement was found to be 112.76 and the lower was -

130.28 for the PASE. The SEM was 46.74 with the SDC of ±129.57. Given that a range of scores for PASE is between 0-400, this represents a large measurement error. The upper 95% limit of agreement for the IPAQ-SF was 4509METS$^{-1minute-1week}$ and lower was -3942METS$^{-1minute-1week}$. This measurement error would seem to be large when considering that the minimum recommended amount of PA in UK Department of Health guidelines is the equivalent of 500METS$^{-1minute-1week}$ (Kaminsky, 2006; Bull & Expert Working Groups, 2010). The large limits of agreement for the IPAQ-SF and PASE are problematic when measuring PA levels in the target population. When looking for change that is not at risk of measurement error, a large change in reported PA levels would be required and is much higher compared to small changes in increasing PA; reducing time spent sedentary that can have positive impacts on health (Proper et al., 2011). Identifying differences in PA between groups also is problematic with the IPAQ-SF and PASE in the target population, as small differences between groups are at risk to measurement error.

In previous OA studies of reliability and measurement error of the two instruments, the repeated measurements have been relatively short: either seven days (Svege et al., 2012; Bolszak et al., 2014) or 9-12 days (Blikman et al., 2013). This thesis assessed a three month time gap between measurements to reflect the reliability and measurement error in measurement periods commonly used in research trials and survey follow ups (Bowling, 2009).

Evaluation of construct validity demonstrated that IPAQ-SF and PASE total scores correlated, suggesting the IPAQ-SF and PASE do measure the same construct. Sub domains of the IPAQ-SF and PASE were also correlated; both instruments

correlated significantly with the physical health component scale of the SF-12 showing discriminant construct validity.

The findings from the systematic review showed responsiveness had not been tested previously for the IPAQ-SF or the PASE. The ability of a self-report PA instrument to detect changes in PA levels is an important measurement property (De Vet et al., 2011). When measuring levels of PA over multiple time points, the PA instrument should have the ability to detect changes in PA behaviour due to interventions or changes in disease progression or health.

Hypothesis testing as outlined by De Vet et al. (2011) was used to determine responsiveness. A sub-sample of responders to the IPAQ-SF and PASE were selected for their likelihood of having increased their PA levels. In practice the sub-sample did not change IPAQ-SF or PASE scores beyond the error of measurement found.

Responsiveness, using hypothesis testing methods (De Vet et al., 2011) could not be evaluated as the sub-sample selected as likely to ‘change PA levels’ did not specify a real change in PA levels on the PASE or IPAQ-SF or in a global assessment of change in their joint problem above the large measurement error. Some indication of the magnitude of change required in the IPAQ-SF and PASE with overall findings suggests that either the PASE or IPAQ-SF can be used to measure PA in the target population. Evaluating change in PA with either measure is challenging due to large measurement error identified with both instruments.

10.5 Strengths and limitations

This section discusses the overall strengths and limitations of each component of the thesis.

10.5.1 Overall strengths and limitations of the thesis

The strengths of the thesis include the size of population available from the MOSAICS study which increases generalisablilty. In the population survey and consultation questionnaires a large number of different variables were measured, including PA. PA was measured in the consultation questionnaire using the IPAQ-SF and PASE which allowed for direct comparisons of their measurement properties. Statistical models could be applied in the large sample in the population survey to describe the associations between joint pain and levels of PA and in the consultation.

Use of the MOSAICS study data also created a number of limitations for the thesis. Measurements were pre-defined prior to commencing the research work within this thesis, and as a consequence, variables and instruments that would have been useful, unfortunately were not available. For example, an objective measurement of PA could have been used as an anchor to compare and test their validity of the instruments in measuring PA. An objective measure, such as an accelerometer, was not used in MOSAICS because of the high cost and practical limitations of collecting data in the large population survey and consultation survey for over 500 participants.

10.5.2 Population survey: Strengths

The population survey was the first survey of PA levels in community dwelling UK adults aged 45 years and over with and without self-reported joint pain. When considering external validity of the findings, the representativeness of the study sample was considered (Sim & Wright, 2000) given that the response rate of the survey was less than 60%.

Selection bias refers to the comparability of the groups that are being studied (Grimes & Schulz, 2002). The population survey limited recruitment bias by inviting all adults registered within eight primary care practices in the North West Midlands and South Cheshire aged 45 years and over using a postal questionnaire. This removed risk of bias of selective encouragement that can occur in other recruitment approaches, such as the telephone.

The study compared levels of PA in adults with and without joint pain and controlled for confounders: gender (Ford et al., 2005), age (Haskell et al., 2007), BMI (Stamatakis et al., 2007) and socio-economic status (Ford et al., 1991). Not adjusting for such factors can cause misinterpretation of results creating spurious relationships between variables (Groenwold et al., 2008); an adjusted model was also implemented to control for confounding.

10.5.3 Population survey: Limitations

Non-response bias was a concern with only 57.1% of survey questionnaires mailed, completed and returned. When non-responders differ in an important way to that of responders in survey research, it creates non-response bias which affects the generalisation of findings (Johnson & Wislar, 2012). Although there is

no gold-standard for an acceptable response rate (Nulty, 2008), in health research a response rate of 60% from a questionnaire could be classed as ‘good’ while 50% is classified as acceptable (Johnson & Wislar, 2012). Compared a previous similar survey study in musculoskeletal research, the response rate in this study was is lower (Thomas et al., 2004). In 2004, a population survey on pain prevalence in an adult population in a similar geographical location reported a response rate of 71.3% (Thomas et al., 2004). A possible explanation for the lower response rate in the MOSAICS population survey was that those with no joint pain potentially did not feel they need to respond. The MOSAICS questionnaire did request that participants should reply whether they had joint pain or not. As the response rate of the MOSAICS population survey was low, to minimise the effect of non-responder bias on findings, comparisons were made with other similar studies (Thomas et al., 2004; Rosemann et al., 2007; Lawrence et al., 2008; Bull & Milton, 2010; Dunlop et al., 2011a; BHF, 2012; White et al., 2013). The joint pain prevelance in this thesis was higher compared to the other studies on pain and OA (Thomas et al., 2004; Lawrence et al., 2008). A possible explanation could include that non-responders had artificially increased the prevalence and the non-responders could have lower joint pain compared with responders.

Another consideration is the presence of joint pain in the MOSAICS survey which was inclusive of pain of any nature in or around the four sites in the past year, representing a large amount of pain over a long period of time in contrast with the other research (Thomas et al., 2004). A comparison between the gender and age of responders and non-responders was conducted; significantly higher percentages of males did not respond as well as non-responders being significantly younger in age compared with responders. A younger age and a

higher percentage of males suggests that the non-responders could have had a lower prevalence of joint pain as opposed to responders, as ageing increases risk of joint pain, and more females than males report joint pain (Loeser, 2013; Neogi & Zhang, 2013).

The threat to internal validity may come from separating two groups of adults aged 45 years and over: those with joint pain in the knee, hip, hand or foot, and those with no reported joint pain. This separation only considers those four sites of pain and not other joint sites that have preponderance in adults aged 45 years and over, such as the lower back, neck or shoulders (Thomas et al., 2004); such limitation means the group classified as not reporting joint pain may still have pain in other joints. This may have had impact when comparisons were made between those with and those without joint pain. It can still be considered that those with self-reported joint pain in any of the four sites are still less likely to be physically active in contrast with those that did not report pain in these sites.

The thesis also has investigated measurement bias. In the MOSAICS population survey, there was some evidence of measurement bias in the STAR questionnaire. An overestimation in levels of PA for adults aged 45 years and over compared with previously reported data on levels of PA in similar-aged, UK adults (BHF, 2012) was evident. There appeared to be an overestimation of PA level in those with self-reported joint pain alongside studies on levels of PA in adults with OA (De Groot et al., 2005; Dunlop et al., 2011a; White et al., 2013). Self-report PA instruments have been shown to overestimate levels of PA whereas objective measures such as accelerometers have not (Helmerhorst et al., 2012). Despite the apparent overestimation, differences found between those who reported joint pain

and no reported joint pain suggest that although the levels of PA may not be representative of true levels due to measurement bias, relative differences may be identified.

10.5.4 Systematic review: Strengths

There were a number of strengths of this systematic review: In Stage A and Stage B included comprehensive searches of a broad selection of relevant electronic databases was used. This was conducted with the aim of maximising the number of relevant articles retrieved for the systematic review and to try to maximise the data available for analysis. Search terms for Stage A and B were selected previously used Cochrane published systematic review of OA and joint pain research (Fransen & McConnell, 2008). Further search terms for measurement properties in patient reported outcome measures were added in the Stage B searches (Terwee et al., 2009). Standardised forms for data extraction and quality assessment were also adapted and used in the systematic review (Terwee et al., 2010; Mokkink et al., 2009). The advantage of using the COSMIN for quality assessment was its uniform definitions and criteria, particularly useful characteristics for undertaking this systematic review as the reviewing team had quite different levels of experience in evaluating measurement properties. Using these standardised forms for data extraction and quality assessment helped to minimise reviewing bias where subjective interpretations during the systematic review process are made based on reviewers' prior knowledge or expectations (Moher et al., 2009). With 52 articles included in Stage B, clear guidance and definitions for terms, such as: reliability, measurement error and the different forms

of validity were needed, therefore, the COSMIN and QAPAQ data extraction checklist were modified to enhance repeatability.

Specially developed forms were constructed for the abstract and full article review phases of Stages A and B. This allowed all reviewers to evaluate each study against the inclusion and exclusion criteria effectively. Using standardised forms allows data extraction and quality assessment processes to be replicable and minimises reviewing bias.

10.5.5 Systematic review: Limitations

A limitation of the systematic review was that title selection could only be conducted by the primary reviewer due to the number of eligible titles retrieved by the search strategy. In Stage A only 50% of the abstracts were double reviewed by a second reviewer. To minimise this impact on the study design, the primary reviewer selected articles after a pilot to minimise reporter bias and error. In the systematic review, the measurement properties were accounted for adults aged 45 years and over with OA or joint pain and community dwelling adults aged 45 years and over separately. It is still not clear if the measurement properties of instruments in community dwelling adults aged 45 years and over is representative for adults aged 45 years and over with joint pain or OA. Consequently the findings in Stage B of the systematic review were separated for popualtions with joint pain and community dwelling adults rather than reported together.

Despite attempts to minimise bias and error there are some limitations to this systematic review. When compared to a previous systematic review (Terwee et al., 2011), a number of instruments were not identified. One explanation for this

could be the exclusion of post-operative joint replacement surgery in the systematic review of this thesis.

10.5.6 Analysis of measurement properties: Strengths

The strengths of the analysis of measurement properties included the direct comparisons of measurement properties of the IPAQ-SF and PASE allowing for a greater understanding of how these two instruments compared. In the MOSAICS consultation questionnaire, the response rate of returned completed questionnaires was 81.9% and of those who completed a baseline questionnaire, 86.8% returned a completed three month follow up questionnaire. This represents a high response rate; therefore the level of generalisablilty was acceptable. The repeated measures study design of the MOSAICS consultation questionnaire allowed for analysis of reliability and responsiveness over time. The three month follow up was a reasonable time interval for evaluating the reliability and responsiveness as OA is a chronic, long term condition and relatively stable over time (Biljlsma et al., 2011) although previous studies have used shorter time intervals (Naal et al., 2009; Svege et al., 2012; Bolszak et al., 2014). A three month time interval also reflects a standard follow up time interval of clinical trials or survey studies (Bowling, 2009). The large sample size also allowed for sub-groups to be selected for the reliability analysis with the sample exceeding the recommended minimum sample of 50 (Giraudeau & Mary, 2001). A minimum number of 50 for a Bland and Altman plot was suggested to be a reasonable number of cases (De Vet et al., 2011). The direct comparisons of the analysis of measurement properties of the IPAQ-SF and PASE were possible as both were included in the MOSAICS consultation questionnaires. This direct comparison of

the two self-report instruments allowed for recommendations for use of the IPAQ-SF or PASE in research. Recent research into measurement properties of the IPAQ-SF and PASE have been in single joint sites rather than focussing on joint pain in the individual (Naal et al., 2009; Terwee et al., 2011; Svege et al., 2012; Bolszak et al., 2014). Assessing measurement properties of PA instruments in adults with joint pain in the hip, knee, hand or foot, or multiple sites of joint pain is also relevant to the NICE guidelines for management of OA, where PA is recommended as a core treatment (NICE, 2014), therefore it was necessary to find an approach to accurately measure PA in adults with joint pain. As a result this allowed the findings in this study to be clinically relevant in adults aged 45 years and over with joint pain, as the MOSAICS population study showed that 69% of adults aged 45 years and over with joint pain had pain in two or more sites. The findings in this study also apply to the clinically defined OA population in the NICE guidelines for management of OA (NICE, 2014).

10.5.7 Analysis of measurement properties: Limitations

A limitation of the analysis of measurement properties was that it was a secondary analysis of MOSAICS data, which did allow for a large dataset but it did not contain a number of parameters which would have allowed for more robust analysis of the instruments' measurement properties. An example of this limitation was in the reliability study where proxy measures of PA levels of change were found to be stable. There were non-significant or small changes in PASE scores, IPAQ-SF scores, pain intensity and health status, suggesting the reliability subsample were stable.

Secondary analysis also limited the analysis of responsiveness with minimal change in PA noted. One explanation for this could have been the IPAQ-SF and PASE were unable to detect change in this subsample, another that the subsample did not change PA behaviour during the time interval. To assess responsiveness in the IPAQ-SF and PASE, two alternative approaches could be used for identifying the sample had changed PA behaviour. The first approach to actually increase PA levels in the sample directly by using supervised PA interventions, then evaluating if the IPAQ-SF or PASE detect the change (van der Bij et al., 2002). A second approach could use an objective measure such as an anchor measurement to select a subsample that had increased levels of PA and compare the changes of IPAQ-SF and PASE scores. Using this approach, responsiveness in the IPAQ-SF and PASE could be assessed in the sample that had increased PA according to the objective measure by identifying if the change in score was greater than the measurement error reported in the IPAQ-SF or PASE (De Vet et al., 2011).

The selection of outcome measures was also a limitation as it would have been preferable for a direct, objective measure of levels in PA to compare the IPAQ-SF and PASE scores. An objective measure of PA, such as an accelerometer would have allowed for an assessment of the IPAQ-SF and PASE in contrast with the objective measure of levels of PA in adults with joint pain or OA using accelerometers to assess construct validity used previously in studies. Svege et al. (2012) used accelerometers to compare the PASE total scores in adults with hip OA and found a low correlation (r=0.30), suggesting the PASE does not have construct validity compared with objective measures. Bolszak et al. (2014) assessed the PASE construct validity against accelerometers in patients with total

knee arthroplasty and found low correlations in males (r=0.45) and in females (r=0.06).

Using an objective measure of PA would have allowed greater accuracy in selecting a reliability sub-sample or responsiveness sub-sample rather than using predictors of being stable or increasing PA; sub-samples could have been selected based on the change or stability of objective PA. For construct validity, comparing IPAQ-SF and PASE scores to accelerometers would have shown the relationship of the IPAQ-SF and PASE to an objective measure of PA levels. If an objective measure was not possible, an alternative method for the analysis of reliability and responsiveness would have been to create a change in PA levels item within the three month survey, similar to global change in perceived health used in general health status instruments (Gandek et al., 1998). From those who reported increasing PA, these individuals could have been selected as a subsample for the responsiveness analysis and those reporting no change could have been selected for reliability analysis. This would be similar to the way that quality of life and health status questionnaires are assessed using a global perceived change anchor method (De Boer et al., 2005). Although this method would have compared self-report levels of PA in those self-reporting increased levels of PA, using an objective instrument as a comparator would have allowed for better assessment of the validity of the IPAQ-SF and PASE. Identifying what would constitute a global change in PA to select a responsiveness subsample would have been difficult. In health-related quality of life measures, this is possible, as a subjective improvement is what the health-related quality of life measures are required to be sensitive to detect. In PA, a subjective change in PA is not as desirable compared to quantifying their PA levels. Asking responders to indicate a

global change in PA levels that is clinically important would be difficult to accurately define, as it depends on the baseline starting level of PA (Haskwell et al., 2007). For example, a magnitude of change that is an important increase of PA level in a sedentary individual would be different to an individual whom is already participating in five days a week of moderate activity lasting at least 30 minutes each day (Powell et al., 2011).

A limitation of the reliability analysis of the IPAQ-SF and leisure activities of the PASE was the changes in scores from baseline to three months were not normally distributed; this meant an ICC was not suitable and so a Spearman's rank correlation test was used. An ICC would have been ideal as it accounts for an individual's variance where Spearman's rank only considers the sample's total variance at the two time points, which then gives associations between the two measurements rather than consistency or agreement.

10.6 Implications and recommendations for future research

OMERACT together with OARSI created a set of outcome measures for clinical trials in OA. OMERACT and OARSI have jointly advocated for the use of physical functioning as a core outcome measure for research in OA, but have not included PA (Dobson et al., 2012; Taylor et al., 2016). Given the importance of PA outcomes for people with OA or joint pain, consideration of PA as a core outcome in OA or joint pain research is warranted. PA outcomes could also be considered by the COMET initiative, which includes consensus on core outcomes for research in OA (Gargon et al., 2014)

There is a large amount of heterogeneity in the instruments used in joint pain and OA studies measuring PA levels, which was demonstrated within Stage A of the

systematic review in this thesis. Twenty three reproducible instruments have previously been used to measure PA in adults aged 45 and over with joint pain or OA. For guidelines on core outcome PA measures for musculoskeletal conditions, evidence is required to identify good quality measurement instruments, and identify instruments of poor quality. There are a number of issues that add complexity to measuring PA in this population. One is the number of PA dimensions that can be measured, for example, total energy expenditure, perceived level of PA or perceived frequency, and duration and intensity of physical activities participated in. Due to these issues, further work is required before PA can be recommended as a core outcome for research in adults with joint pain or OA. There would firstly need to be some work to develop consensus, with relevant stakeholders within the joint pain or OA research community on what domains of PA are important to include for research in joint pain or OA (Williamson et al., 2012).

The other key issue is identifying instruments with acceptable measurement properties. This thesis has shown the limitations of the STAR, PASE and IPAQ-SF in their measurement of PA levels in the target population.

The findings from this thesis can be taken forward for use in future research. With a greater emphasis on PA as a core treatment for OA (NICE, 2008; 2014), more research could focus on targeting those with joint pain who are at risk of low levels of PA in primary care. Based on the findings from the population survey, it appears that those with joint pain are at greater risk of lower PA levels, compared to those with no self-reported joint pain. Therefore, further research on those with joint pain should focus upon the barriers and attitudes to PA, and target these individuals to

increase their PA levels. Based on the findings of the thesis, the STAR appeared to overestimate levels of PA in the MOSAICS population survey. In extensive population surveys, larger multiple item questionnaires such as the IPAQ-SF and PASE are often not possible. This could be due to a restriction in the number of pages a population survey questionnaire contains. Shorter surveys are desirable, as they have higher average response rates to longer surveys (Edwards et al., 2002). As the STAR overestimated PA, compared to other self-reporting instruments, it would not seem to be an appropriate instrument.

If future research is conducted which is only interested in measuring a summary of PA in a joint pain or OA population, it is apparent that the single item scales identified in the systematic review may supply the same information as longer instruments, such as the IPAQ-SF and PASE. Single scale items can rank individuals into different levels of physical activity, which will still allow research questions to be answered. Single item scales were evaluated in the systematic review, including: the Activity Rating Scale (ARS), the Tegner scale, and the University College of Los Angles Activity (UCLAA) scale (Naal et al., 2009). The advantage of using such single scale items compared to using the STAR, is that the scores have been shown to correlate with the IPAQ-SF, and to be reliable in adults aged 45 years and over with total joint replacement surgery (Naal et al., 2009).

Further assessment of the single scale items would be required before they can be recommended ahead of the STAR, but they do represent alternative instruments for population-level surveys in adults aged 45 years and over with joint pain. The assessment of these single-item scales should initially include working with patient stakeholder groups. This can be conducted to help determine whether the

instruments are acceptable for use in the target population. This can be achieved by using cognitive interviewing of the target population to evaluate acceptability (Heesch et al., 2011; 2012). The single item scale instrument should also be further tested for reliability, measurement error, construct validity and responsiveness.

Single scale items may not be viable as measures in all research, as they can only rank PA and cannot provide greater detail for duration, frequency and intensity of different physical activities. Because of this, other measures of PA are also required. The systematic review and analysis of measurement properties showed the low quality of measurement properties in the PASE and IPAQ-SF, and demonstrated that objective measures should be recommended where possible, in order to determine levels of PA in research with adults aged 45 years and over with joint pain or OA. This may not always be possible at the current time, as there are limitations of using objective measures of PA in studies of larger sample sizes or population surveys, in terms of high cost, and practicality of collecting objective measurements.

Measuring PA in this population is important when testing interventions that aim to increase levels of PA which can lead to improvements in pain and physical function outcomes, as well as general physical and mental health outcomes.

When using the IPAQ-SF to measure PA in populations, data can be heavily skewed, as was found in this thesis and also in other studies (Craig et al., 2003; Mader et al., 2006). Because of these limitations of poor completion rates, the floor effect of total scores, and poor measurement properties in measurement error, the IPAQ-SF is not recommended for use in adults aged 45 and over with

joint pain. It should be considered that when using either the PASE or IPAQ-SF, a large amount of measurement error is likely in such adults.

Further evaluation of the PASE and IPAQ-SF may be considered in future research. Evaluating the reliability and responsiveness of both IPAQ-SF and PASE using objective measures such as activity monitors, would provide additional evidence for precision in reliability and responsiveness. Using activity monitors as the standard measurement of a reliability and responsiveness indicator, would allow for greater understanding of both the IPAQ-SF and PASE's measurement properties in the target population.

With the IPAQ-SF and PASE deemed inappropriate for the target population, new measures could be developed. The findings from the systematic review suggest that when measuring PA, in order to establish a single overall indicator of the PA level in the target population, single item scales could be appropriate as they appear to have similar measurement properties to other instruments. Further, investigation of the single item scales identified in the systematic review should be considered. The first investigation should be to evaluate their content and face validity by interviewing adults aged 45 years and over with joint pain, to test their acceptability and feasibility. Further evaluations could also be conducted to examine validity, reliability, measurement error and responsiveness to the target population. Direct comparison of these measurement properties should then be made with objective measures of PA levels, in order to compare overall performance.

Another consideration in measuring PA behaviour is the development of a new specific measure for adult populations with OA or joint pain. The self-reporting PA

instruments identified in this thesis have all been developed using an expert and evidence-based approach. For example, the IPAQ-SF was developed by an international collaboration of expert panels including clinicians, sports scientists and researchers (Craig et al., 2003). The PASE was developed by a combination of a systematic review of relevant PA literature, and consultation with PA experts (Washburn et al., 1993).

An established method, used previously in different areas in musculoskeletal populations, involves the use of patient populations as participants in development of the items (Food and Drug Administration (FDA), 2009). Patient populations as research participants should be included as key members and be involved in the development of any new instrument. The FDA guidance of the eight steps of design in relation to content validity should also be followed when developing a new instrument (FDA, 2009). Guidance on involving patients in the decision-making process of making modifications during development of instruments are also provided by the FDA (FDA, 2009).

This approach does not appear to have been used in any of the PA measures for OA or joint pain populations identified in this thesis. Cognitive interview methods could be used to examine the limitations of the PASE and IPAQ-SF (Heesch et al., 2011; 2012). Patients and the public as partners includes them taking an active role within the work of the research, through the whole research process and involvement in key decision-making throughout (Staniszewska et al., 2012). This can include the development, evaluation and modification stages of new self-report instruments. Patients and public partners can decide on the importance and relevance of PA measurement in OA and joint pain populations, interpret

qualitative data on the domains of interest for the self-report PA instrument, decide the number of items that should be included, and the final wording of items (Blackburn et al., 2016; Haywood et al., 2016). Patients as study participants are recruited to the research itself and generate the data for analysis (Staniszewska et al., 2012).

The aim of involving patients and the public as both collaborators and participants in the development of any new self-report PA instrument is to increase relevance and acceptability of the instrument (Staniszewska et al., 2012). The purpose of this strategy is to reduce the discrepancies observed in previous self-report outcomes and patients' own outcomes (Haywood et al., 2016).

One approach to working with patients and the public as partners is described by Blackburn et al. in the development of patient-reported quality indicators of OA (Blackburn et al., 2016). There were three stages; Stage 1 develops the questionnaire, understanding the aims of the measure in the target population, identifying relevant domains, then developing items; Stage 2 makes comparisons with other instruments, Stage 3 reviews the finalised version, using field testing with patient and public partners. This systematic approach allows for clarity in reporting the role of the patients as partners in the research (Blackburn et al., 2016). A Research User Group such as the one at Keele University (Jinks et al., 2013) can support such approaches to patient and public involvement and engagement.

Development of a new self-report PA instrument including patients with joint pain or OA, along with clinicians and researchers as participants, could help identify the key domains and subdomains that are important. Item generation should also

be conducted with the involvement of all stakeholders including patients, clinicians, and researchers as equal study participants. Consideration should also be given to the structure of the new instrument; there is an opportunity not only to give a summary score, but also one for separate domains.

A new instrument as described previously would differ from other instruments, as it would be developed specifically for a joint pain or OA population, using patient stakeholders as collaborators in the instrument's development. The advantage of using a patient-derived approach is that the items would be more appropriate and acceptable, which may possibly avoid the problem of misunderstanding that has been reported in previous instruments (Haywood et al., 2010; Heesch et al., 2011; 2012).

The systematic review in this thesis showed that the majority of self-reporting PA instruments are only assessed for their reliability, measurement error, construct/criterion validity, and occasionally responsiveness. When piloting and evaluating a new self-reporting PA instrument for a joint pain or OA population, testing of a wider number of measurement properties including internal consistency, face and content validity, structural validity of domains and sub domains, and cross-cultural validity would be recommended (Mokkink et al., 2010). Acceptability and feasibility should also be assessed. For example, testing the completion rates and burden to complete are also important in assessing any new instrument (Streiner & Norman, 2008). This wider evaluation would allow identification of potential limitations of the new self-reporting instrument, which can be re-visited.

It is not clear at this time which objective measure of PA in the target population is the most suitable. Future research on the measurement properties of objective measures of PA in this population should be considered. The ease of use of these devices for adults aged 45 years and over with joint pain should also be evaluated, as devices that require user-calibration may be difficult for adults aged 45 years and over with joint pain. An example of an objective device being evaluated is the Sensewear Pro^3 activity monitor which has been shown to over-estimate energy expenditure in adults with hip OA compared with indirect calorimetry (Hermann et al., 2014). A single measurement device should be used to create a uniform method of reporting levels of PA across all joint pain and OA research reducing reporting bias when combining data in literature syntheses.

10.7 Final thesis conclusion

This thesis has demonstrated how measuring PA using self-report methods in the target population is important in clinical research. By using self-report instruments to measure PA, it has been identified that at a population level adults aged 45 years and over with reported joint pain are less active than those with no reported joint pain. It has also shown that higher levels of PA in adults aged 45 years and over with joint pain is associated with better health-related quality of life and that those with painful joints in both upper and lower limbs have lower levels of PA compared to those with joint pain in the lower limbs only. The thesis has also detailed the implications of using self-report measures of PA in adults aged 45 years and over with joint pain and OA and recommends, where possible, to use a measure of PA consistently. Based upon the findings within this thesis, neither the IPAQ-SF or PASE are deemed suitable self-report measures of PA in adults with

joint pain, the limitations of these instruments in completion rates, measurement error, reliability and validity should be considered.

11 References

Achten, J. & Jeukendrup, A.E. (2003). Heart rate monitoring. *Sports medicine.* 33 (7): 517-538.

Adams, S.A., Matthews, C.E., Ebbeling, C.B., Moore, C.G., Cunningham, J.E., Fulton, J. & Hebert, J.R. (2005). The effect of social desirability and social approval on self-reports of physical activity. *American Journal of Epidemiology.* 161 (4): 389-398.

Adamson, J., Ebrahim, S., Dieppe, P. & Hunt, K. (2006). Prevalence and risk factors for joint pain among men and women in the West of Scotland Twenty-07 study. Annals of the Rheumatic Diseases. 65 (4): 520-524.

Ageberg, E., Engstrom, G., Gerhardsson de Verdier, M., Rollof, J., Roos, E.M. & Lohmander, L.S. (2012). Effect of leisure time physical activity on severe knee or hip osteoarthritis leading to total joint replacement: a population-based prospective cohort study. BMC Musculoskeletal Disorders. 13 : 73.

Ainsworth, B.E. (2009). How do I measure physical activity in my patients? Questionnaires and objective methods. *British journal of sports medicine.* 43 (1): 6-9.

Allender, S., Foster, C., Scarborough, P. & Rayner, M. (2007). The burden of physical activity-related ill health in the UK. *Journal of epidemiology and community health.* 61 (4): 344-348.

Altman, D.G. (1990). *Practical statistics for medical research.* London: Chapman and Hall.

Altman, R., Alarcon, G., Appelrouth, D., Bloch, D., Borenstein, D., Brandt, K., Brown, C., Cooke, T., Daniel, W. & Feldman, D. (1991). The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. *Arthritis & Rheumatism.* 34 (5): 505-514.

Altman, R., Alarcon, G., Appelrouth, D., Bloch, D., Borenstein, D., Brandt, K., Brown, C., Cooke, T., Daniel, W. & Gray, R. (1990). The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hand. *Arthritis & Rheumatism.* 33 (11): 1601-1610.

Altman, R., Asch, E., Bloch, D., Bole, G., Borenstein, D., Brandt, K., Christy, W., Cooke, T., Greenwald, R. & Hochberg, M. (1986). Development of criteria for the classification and reporting of osteoarthritis: classification of osteoarthritis of the knee. *Arthritis & Rheumatism.* 29 (8): 1039-1049.

Analay, Y., Ozcan, E., Karan, A., Diracoglu, D. & Aydin, R. (2003). The effectiveness of intensive group exercise on patients with ankylosing spondylitis. *Clinical rehabilitation.* 17 (6): 631-636.

Anderson, D.D., Chubinskaya, S., Guilak, F., Martin, J.A., Oegema, T.R., Olson, S.A. & Buckwalter, J.A. (2011). Post-traumatic osteoarthritis: Improved understanding and opportunities for early intervention. *Journal of Orthopaedic Research.* 29 (6): 802-809.

Arden, N. & Nevitt, M.C. (2006). Osteoarthritis: epidemiology. *Best practice & research. Clinical Rheumatology.* 20 (1): 3-25.

Armstrong, T., Bauman, A.E. & Davies, J. (2000). *Physical activity patterns of Australian adults: results of the 1999 National Physical Activity Survey.* Australian Institute of Health and Welfare.

Arokoski, M.H., Arokoski, J.P., Haara, M., Kankaanpaa, M., Vesterinen, M., Niemitukia, L.H. & Helminen, H.J. (2002). Hip muscle strength and muscle cross sectional area in men with and without hip osteoarthritis. *The Journal of rheumatology.* 29 (10): 2185-2195.

Arthritis Research UK (ARUK) (2013) *Osteoarthritis in general practice: Data and perspectives.* Available at: *http://www.arthritisresearchuk.org/arthritis-information/data-and statistics/~/media/EFAEFCE432734F3AA5FB1C64329E02D1.ashx* (Accessed: 9th June 2014)

Ashley, A., Lloyd, A., Lamb, S. & Bartlett, H. (2001). Is health-related quality of life a suitable outcome measure for evaluating health promotion programmes? *Nursing Times Research.* 6 (3): 671-678.

Aspden, R.M. (2011). Obesity punches above its weight in osteoarthritis. *Nature Reviews Rheumatology.* 7 (1): 65-68.

Baecke, J.A., Burema, J. & Frijters, J.E. (1982). A short questionnaire for the measurement of habitual physical activity in epidemiological studies. *The American Journal of Clinical Nutrition.* 36 (5): 936-942.

Balady, G.J., Williams, M.A., Ades, P.A., Bittner, V., Comoss, P., Foody, J.M., Franklin, B., Sanderson, B., Southard, D., American Heart Association Exercise, Cardiac Rehabilitation, and Prevention Committee, the Council on Clinical Cardiology, American Heart Association Council on Cardiovascular Nursing, American Heart Association Council on Epidemiology and Prevention, American Heart Association Council on Nutrition, Physical Activity, and Metabolism & American Association of Cardiovascular and Pulmonary Rehabilitation (2007).

Core components of cardiac rehabilitation/secondary prevention programs: 2007 update: a scientific statement from the American Heart Association Exercise, Cardiac Rehabilitation, and Prevention Committee, the Council on Clinical Cardiology; the Councils on Cardiovascular Nursing, Epidemiology and Prevention, and Nutrition, Physical Activity, and Metabolism; and the American Association of Cardiovascular and Pulmonary Rehabilitation. *Circulation.* 115 (20): 2675-2682.

Barbour, K.E., Hootman, J.M., Helmick, C.G., Murphy, L.B., Theis, K.A., Schwartz, T.A., Kalsbeek, W.D., Renner, J.B. & Jordan, J.M. (2014). Meeting physical activity guidelines and the risk of incident knee osteoarthritis: a population-based prospective cohort study. *Arthritis care & research*. 66 (1): 139-146.

Bartels, E., Lund, H., Hagen, K., Dagfinrud, H., Christensen, R. & Danneskiold-Samsøe, B. (2007). Aquatic exercise for the treatment of knee and hip osteoarthritis. *Cochrane Database Syst Rev.* 4.

Batsis, J.A., Zbehlik, A.J., Barre, L.K., Mackenzie, T.A. & Bartels, S.J. (2014). The impact of waist circumference on function and physical activity in older adults: longitudinal observational data from the osteoarthritis initiative. *Nutrition Journal.* 13:81.

Batsis, J.A., Zbehlik, A.J., Scherer, E.A., Barre, L.K. & Bartels, S.J. (2015). Normal Weight with Central Obesity, Physical Activity, and Functional Decline: Data from the Osteoarthritis Initiative. *Journal of the American Geriatrics Society.* 63 (8):1552-1560.

Bauman, S., Williams, D., Petruccelli, D., Elliott, W. & de Beer, J. (2007). Physical activity after total joint replacement: a cross-sectional survey. *Clinical journal of*

sport medicine: official journal of the Canadian Academy of Sport Medicine. 17 (2): 104-108.

Bautch, J.C., Malone, D.G. & Vailas, A.C. (1997). Effects of exercise on knee joints with osteoarthritis: a pilot study of biologic markers. *Arthritis & Rheumatism.* 10 (1): 48-55.

Bellamy, N., Kirwan, J., Boers, M., Brooks, P., Strand, V., Tugwell, P., Altman, R., Brandt, K., Dougados, M. & Lequesne, M. (1997). Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. *J Rheumatol.* 24 (4): 799-802.

Bennell, K.L., Ahamed, Y., Bryant, C., Jull, G., Hunt, M.A., Kenardy, J., Forbes, A., Harris, A., Nicholas, M., Metcalf, B., Egerton, T. & Keefe, F.J. (2012). A physiotherapist-delivered integrated exercise and pain coping skills training intervention for individuals with knee osteoarthritis: a randomised controlled trial protocol. *BMC Musculoskeletal Disorders.* 13 : 129.

Bennell, K.L., Egerton, T., Bills, C., Gale, J., Kolt, G.S., Bunker, S.J., Hunter, D.J., Brand, C.A., Forbes, A., Harris, A. & Hinman, R.S. (2012a). Addition of telephone coaching to a physiotherapist-delivered physical activity program in people with knee osteoarthritis: a randomised controlled trial protocol. *BMC Musculoskeletal Disorders.* 13 : 24

Bennell, K.L., Hinman, R.S., Crossley, K.M., Metcalf, B.R., Buchbinder, R., Green, S. & McColl, G. (2004). Is the Human Activity Profile a useful measure in people with knee osteoarthritis? *Journal of Rehabilitation Research & Development*. 41 (4): 621-30.

Bennell, K.L., Hunt, M.A., Wrigley, T.V., Lim, B. & Hinman, R.S. (2008). Role of muscle in the genesis and management of knee osteoarthritis. *Rheumatic Disease Clinics of North America.* 34 (3): 731-754.

Berenbaum, F., Eymard, F. & Houard, X. (2013). Osteoarthritis, inflammation and obesity. *Current opinion in rheumatology.* 25 (1): 114-118.

Bijlsma, J.W., Berenbaum, F. & Lafeber, F.P. (2011). Osteoarthritis: an update with relevance for clinical practice. *The Lancet.* 377 (9783): 2115-2126.

Bilek, L.D., Venema, D.M., Camp, K.L., Lyden, E.R. & Meza, J.L. (2005). Evaluation of the human activity profile for use with persons with arthritis. *Arthritis & Rheumatism*. 53 (5): 756-63.

Blikman, T., Stevens, M., Bulstra, S.K., van den Akker-Scheek, I. & Reininga, I.H.F. (2013). Reliability and validity of the Dutch version of the International Physical Activity Questionnaire in patients after total hip arthroplasty or total knee arthroplasty. *Journal of Orthopaedic & Sports Physical Therapy.* 43 (9): 650-659.

Bize, R., Johnson, J.A. & Plotnikoff, R.C. (2007). Physical activity level and health-related quality of life in the general adult population: a systematic review. *Preventive medicine.* 45 (6): 401-415.

Blagojevic, M., Jinks, C., Jeffery, A. & Jordan, K. (2010). Risk factors for onset of osteoarthritis of the knee in older adults: a systematic review and meta-analysis. *Osteoarthritis and cartilage.* 18 (1): 24-33.

Blair, S.N. & Brodney, S. (1999). Effects of physical inactivity and obesity on morbidity and mortality: current evidence and research issues. *Medicine and science in sports and exercise.* 31: S646-S662.

Blair, S.N., Horton, E., Leon, A.S., Lee, I.M., Drinkwater, B.L., Dishman, R.K., Mackey, M. & Kienholz, M.L. (1996). Physical activity, nutrition, and chronic disease. *Medicine and science in sports and exercise.* 28 (3): 335-349.

Blair, S.N., Kohl, H.W., III, Gordon, N.F. & Paffenbarger Jr, R.S. (1992). How much physical activity is good for health? *Annual Review of Public Health.* 13 (1): 99-126.

Blair SN, Kohl H.W., III, Paffenbarger RS, Jr, Clark DG, Cooper KH, Gibbons LW (1989). Physical fitness and all-cause mortality: A prospective study of healthy men and women. *JAMA.* 262 (17): 2395-2401.

Blair, S.N. & Morris, J.N. (2009). Healthy hearts—and the universal benefits of being physically active: physical activity and health. *Annals of Epidemiology.* 19 (4): 253-256.

Bland, J.M. (2000). *An introduction to medical statistics.* Oxford University Press.

Bland, J.M. & Altman, D.G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1 (8476): 307-310.

Bland, J.M. & Altman, D.G. (1999). Measuring agreement in method comparison studies. *Statistical methods in medical research.* 8 (2): 135-160.

Blikman, T., Stevens, M., Bulstra, S.K., van den Akker-Scheek, I. & Reininga, I.H. (2013). Reliability and Validity of the Dutch Version of the International Physical Activity Questionnaire in Patients After Total Hip Arthroplasty or Total Knee Arthroplasty. *Journal of orthopaedic & sports physical therapy.* 43 (9): 650-659.

Boers, M., Brooks, P., Simon, L., Strand, V. & Tugwell, P. (2005). OMERACT: an international initiative to improve outcome measurement in rheumatology. *Clinical and experimental rheumatology.* 23 (5): S10.

Bolszak, S., Casartelli, N.C., Impellizzeri, F.M. & Maffiuletti, N.A. (2014). Validity and reproducibility of the Physical Activity Scale for the Elderly (PASE) questionnaire for the measurement of the physical activity level in patients after total knee arthroplasty. *BMC musculoskeletal disorders.* 15 : 46-2474-15-46.

Bondeson, J., Wainwright, S.D., Lauder, S., Amos, N. & Hughes, C.E. (2006). The role of synovial macrophages and macrophage-produced cytokines in driving aggrecanases, matrix metalloproteinases, and other destructive and inflammatory responses in osteoarthritis. *Arthritis research & therapy.* 8 (6): R187.

Bonomi, A.G., Plasqui, G., Goris, A.H. & Westerterp, K.R. (2010). Estimation of Free-Living Energy Expenditure Using a Novel Activity Monitor Designed to Minimize Obtrusiveness. *Obesity.* 18 (9): 1845-1851.

Bonnet, C.S. & Walsh, D.A. (2005). Osteoarthritis, angiogenesis and inflammation. *Rheumatology (Oxford, England).* 44 (1): 7-16.

Booth, M.L., Ainsworth, B.E., Pratt, M., Ekelund, U., Yngve, A., Sallis, J.F. & Oja, P. (2003). International physical activity questionnaire: 12-country reliability and validity. *Med sci sports Exerc.* 195 (9131/03): 3508-1381.

Börjesson, M., Robertson, E., Weidenhielm, L., Mattsson, E. & Olsson, E. (1996). Physiotherapy in knee osteoarthrosis: effect on pain and walking. *Physiotherapy Research International.* 1 (2): 89-97.

Bossen, D., Buskermolen, M., Veenhof, C., de Bakker, D. & Dekker, J. (2013). Adherence to a web-based physical activity intervention for patients with knee and/or hip osteoarthritis: a mixed method study. *Journal of Medical Internet Research.* 15 (10): e223.

Bossen, D., Veenhof, C., Van Beek, K.E., Spreeuwenberg, P.M., Dekker, J. & De Bakker, D.H. (2013). Effectiveness of a web-based physical activity intervention in patients with knee and/or hip osteoarthritis: randomized controlled trial. *Journal of Medical Internet Research.* 15 (11): e257.

Bot, S.D., Terwee, C.B., van der Windt, D.A., Feleus, A., Bierma-Zeinstra, S.M., Knol, D.L., Bouter, L.M. & Dekker, J. (2004). Internal consistency and validity of a new physical workload questionnaire. *Occupational and environmental medicine.* 61 (12): 980-986.

Bouchard, C., Blair, S.N. & Haskell, W.L. (2012). *Physical activity and health. (*2nd Ed). Champaign, United States: Human Kinetics.

Boule, N.G., Haddad, E., Kenny, G.P., Wells, G.A. & Sigal, R.J. (2001). Effects of exercise on glycemic control and body mass in type 2 diabetes mellitus: a meta-analysis of controlled clinical trials. *Jama.* 286 (10): 1218-1227.

Bowling, A. (2009). *Research methods in health.* Open University Press Maidenhead, United Kingdom.

Brand, R., Schlicht, W., Grossmann, K. & Duhnsen, R. (2006). Effects of a physical exercise intervention on employees' perceptions of quality of life: a randomized controlled trial. *Sozial-und Präventivmedizin.* 51 (1): 14-23.

Brazier, J., Jones, N. & Kind, P. (1993). Testing the validity of the Euroqol and comparing it with the SF-36 health survey questionnaire. *Quality of Life Research.* 2 (3): 169-180.

Brouwer, C.N., Schilder, A.G., van Stel, H.F., Rovers, M.M., Veenhoven, R.H., Grobbee, D.E., Sanders, E.A. & Maillé, A.R. (2007). Reliability and validity of functional health status and health-related quality of life questionnaires in children with recurrent acute otitis media. *Quality of Life Research.* 16 (8): 1357-1373.

Brown, D.W., Balluz, L.S., Heath, G.W., Moriarty, D.G., Ford, E.S., Giles, W.H. & Mokdad, A.H. (2003). Associations between recommended levels of physical activity and health-related quality of life findings from the 2001 Behavioral Risk Factor Surveillance System (BRFSS) survey. *Preventive medicine.* 37 (5): 520-528.

Brown, W.J., Burton, N.W., Marshall, A.L. & Miller, Y.D. (2008). Reliability and validity of a modified self-administered version of the Active Australia physical activity survey in a sample of mid-age women. *Australian & New Zealand Journal of Public Health*. 32 (6): 535-41.

Brown, W.J., Trost, S.G., Bauman, A., Mummery, K. & Owen, N. (2004). Test-retest reliability of four physical activity measures used in population surveys. *Journal of Science & Medicine in Sport*. 7 (2): 205-15.

Brown, W.J., Williams, L., Ford, J.H., Ball, K. & Dobson, A.J. (2005). Identifying the energy gap: magnitude and determinants of 5-year weight gain in midage women. *Obesity research.* 13 (8): 1431-1441.

British Heart Foundation (BHF) Health Promotion Research Group (2012). *Physical activity statistics 2012.* London, UK: British Heart Foundation.

Bryman, A. (2012). *Social research methods.* Oxford university press.

Buchanan, W.W., Kean, W.F. & Kean, R. (2003). History and current status of osteoarthritis in the population. *Inflammopharmacology.* 11 (4-6): 301-316.

Bull, F. & the Expert Working Groups (2010). *Physical Activity Guidelines in the U.K.: Review and Recommendations. School of Sport, Exercise and Health Sciences.* Loughborough University, United Kingdom: Department of Health.

Bull, F.C. & Milton, K.E. (2010). A process evaluation of a" physical activity pathway" in the primary care setting. *BMC Public Health.* 10 (1): 1-9.

Buckwalter, J.A. & Lane, N.E. (1997). Athletics and osteoarthritis. *The American Journal of Sports Medicine.* 25 (6): 873-881.

Busch, A.J., Schachter, C.L., Overend, T.J., Peloso, P.M. & Barber, K.A. (2008). Exercise for fibromyalgia: a systematic review. *The Journal of rheumatology.* 35 (6): 1130-1144.

Cahue, S., Dunlop, D., Hayes, K., Song, J., Torres, L. & Sharma, L. (2004). Varus–valgus alignment in the progression of patellofemoral osteoarthritis. *Arthritis & Rheumatism.* 50 (7): 2184-2190.

Carek, P.J., Laibstain, S.E. & Carek, S.M. (2011). Exercise for the treatment of depression and anxiety. *The International Journal of Psychiatry in Medicine.* 41 (1): 15-28.

Carville, S.F., Arendt-Nielsen, S., Bliddal, H., Blotman, F., Branco, J.C., Buskila, D., Da Silva, J.A., Danneskiold-Samsoe, B., Dincer, F., Henriksson, C., Henriksson, K.G., Kosek, E., Longley, K., McCarthy, G.M., Perrot, S., Puszczewicz, M., Sarzi-Puttini, P., Silman, A., Spath, M., Choy, E.H. & EULAR

(2008). EULAR evidence-based recommendations for the management of fibromyalgia syndrome. *Annals of the Rheumatic Diseases.* 67 (4): 536-541.

Casartelli, N.C., Bolszak, S., Impellizzeri, F.M. & Maffiuletti, N.A. (2015). Reproducibility and validity of the physical activity scale for the elderly (PASE) questionnaire in patients after total hip arthroplasty. *Physical Therapy.* 95 (1): 86-94.

Caspersen, C.J., Bloemberg, B.P., Saris, W.H., Merritt, R.K. & Kromhout, D. (1991). The prevalence of selected physical activities and their relation with coronary heart disease risk factors in elderly men: the Zutphen Study, 1985. *American Journal of Epidemiology.* 133 (11): 1078-1092.

Caspersen, C.J., Powell, K.E. & Christenson, G.M. (1985). Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. *Public health reports.* 100 (2): 126-131.

Ceesay, S.M., Prentice, A.M., Day, K.C., Murgatroyd, P.R., Goldberg, G.R., Scott, W. & Spurr, G. (1989). The use of heart rate monitoring in the estimation of energy expenditure: a validation study using indirect whole-body calorimetry. *British Journal of Nutrition.* 61 (2): 175-186.

Cerejo, R., Dunlop, D.D., Cahue, S., Channin, D., Song, J. & Sharma, L. (2002). The influence of alignment on risk of knee osteoarthritis progression according to baseline stage of disease. *Arthritis & Rheumatism.* 46 (10): 2632-2636.

Cerin, E., Barnett, A., Cheung, M., Sit C, H.P., Macfarlane, D.J. & Chan, W. (2012). Reliability and validity of the IPAQ-L in a sample of Hong Kong urban older adults: does neighborhood of residence matter?. *Journal of Aging & Physical Activity.* 20 (4): 402-420.

Chaipinyo, K. & Karoonsupcharoen, O. (2009). No difference between home-based strength training and home-based balance training on pain in patients with knee osteoarthritis: a randomised trial. *Australian Journal of Physiotherapy.* 55 (1): 25-30.

Chan, C.B., Ryan, D.A. & Tudor-Locke, C. (2006). Relationship between objective measures of physical activity and weather: a longitudinal study. *The international journal of behavioral nutrition and physical activity.* 3: 21.

Chen, A., Gupte, C., Akhtar, K., Smith, P. & Cobb, J. (2012). The global economic cost of osteoarthritis: how the UK compares. *Arthritis.* 2012.

Chen, K.Y. & Bassett, D.R. (2005). The technology of accelerometry-based activity monitors: current and future. *Medicine and science in sports and exercise.* 37 (11): S490.

Chinapaw, M.J., Mokkink, L.B., van Poppel, M.N., van Mechelen, W. & Terwee, C.B. (2010). Physical activity questionnaires for youth: a systematic review of measurement properties. *Sports medicine.* 40 (7): 539-563.

Chmelo, E., Nicklas, B., Davis, C., Miller, G.D., Legault, C. & Messier, S. (2013). Physical activity and physical function in older adults with knee osteoarthritis. *Journal of Physical Activity & Health.* 10 (6): 777-783.

Chou, R. & Huffman, L.H. (2007). Nonpharmacologic therapies for acute and chronic low back pain: a review of the evidence for an American Pain Society/American College of Physicians clinical practice guideline. *Annals of Internal Medicine.* 147 (7): 492-504.

Christensen, K., Doblhammer, G., Rau, R. & Vaupel, J.W. (2009). Ageing populations: the challenges ahead. *The Lancet.* 374 (9696): 1196-1208.

Cochrane, T., Davey, R.C. & Matthews Edwards, S.M. (2005). Randomised controlled trial of the cost-effectiveness of water-based therapy for lower limb osteoarthritis. *Health technology assessment.* 9 (31): iii-iv, ix-xi, 1-114.

Coggon, D., Croft, P., Kellingray, S., Barrett, D., McLaren, M. & Cooper, C. (2000). Occupational physical activities and osteoarthritis of the knee. *Arthritis & Rheumatism.* 43 (7): 1443-1449.

Colbert, L.H., Matthews, C.E., Havighurst, T.C., Kim, K. & Schoeller, D.A. (2011). Comparative validity of physical activity measures in older adults. *Medicine and science in sports and exercise.* 43 (5): 867-876.

Cohen, J. (1988). Set correlation and contingency tables. *Applied Psychological Measurement*. 12 (4): 425-434.

Conaghan, P.G. (2002). Update on osteoarthritis part 1: current concepts and the relation to exercise. *British journal of sports medicine.* 36 (5): 330-333.

Conaghan, P.G., Porcheret, M., Kingsbury, S.R., Gammon, A., Soni, A., Hurley, M., Rayman, M.P., Barlow, J., Hull, R.G. & Cumming, J. (2014). Impact and therapy of osteoarthritis: the Arthritis Care OA Nation 2012 survey. *Clinical rheumatology.* : 1-8.

Cooper, C., Inskip, H., Croft, P., Campbell, L., Smith, G., McLaren, M. & Coggon, D. (1998). Individual risk factors for hip osteoarthritis: obesity, hip injury, and physical activity. *American Journal of Epidemiology.* 147 (6): 516-522.

Corbin, C.B., Pangrazi, R.P. & Franks, B.D. (2000). Definitions: Health, Fitness, and Physical Activity. *President's Council on Physical Fitness and Sports Research Digest.*

Corder, K., Brage, S. & Ekelund, U. (2007). Accelerometers and pedometers: methodology and clinical application. *Current opinion in clinical nutrition and metabolic care.* 10 (5): 597-603.

Coughlin, S.S. (1990). Recall bias in epidemiologic studies. *Journal of clinical epidemiology.* 43 (1): 87-91.

Craft, L.L. & Perna, F.M. (2004). The Benefits of Exercise for the Clinically Depressed. *Primary care companion to the Journal of clinical psychiatry.* 6 (3): 104-111.

Craig, C.L., Lambert, E.V., Inoue, S., Alkandari, J.R., Leetongin, G. & Kahlmeier, S. (2012). The pandemic of physical inactivity: global action for public health. *The Lancet.* 380 (9838): 294-305.

Craig, C.L., Marshall, A.L., Sjostrom, M., Bauman, A.E., Booth, M.L., Ainsworth, B.E., Pratt, M., Ekelund, U., Yngve, A., Sallis, J.F. & Oja, P. (2003). International PA questionnaire: 12-country reliability and validity. *Medicine and Science in Sports and Exercise*. 35 (8): 1381-95.

Creamer, P. & Hochberg, M.C. (1998). The relationship between psychosocial variables and pain reporting in osteoarthritis of the knee. *Arthritis & Rheumatism.* 11 (1): 60-65.

Croft, P., Blyth, F.M. & van der Windt, D. (2010). *Chronic pain epidemiology: from aetiology to public health.* Oxford University Press.

Croft, P., Coggon, D., Cruddas, M. & Cooper, C. (1992). Osteoarthritis of the hip: an occupational disease in farmers. *BMJ (Clinical research ed.).* 304 (6837): 1269-1272.

Cushnaghan, J. & Dieppe, P. (1991). Study of 500 patients with limb joint osteoarthritis. I. Analysis by age, sex, and distribution of symptomatic joint sites. *Annals of the Rheumatic Diseases.* 50 (1): 8-13.

Dagfinrud, H., Kvien, T.K. & Hagen, K.B. (2005). The Cochrane review of physiotherapy interventions for ankylosing spondylitis. *The Journal of rheumatology.* 32 (10): 1899-1906.

Daskapan, A., Tuzun, E.H. & Eker, L. (2005). Relationship between physical activity level and health related quality of life among university students. *Saudi medical journal.* 26 (6): 1026-1028.

Das, P. & Horton, R. (2012). Rethinking our approach to physical activity. *The Lancet.* 380 (9838): 189-190.

Davidson, M. & de Morton, N. (2007). A systematic review of the Human Activity Profile. *Clinical rehabilitation.* 21 (2): 151-162.

Dawson, J., Juszczak, E., Thorogood, M., Foster, C., Marks, S., Dodd, C. & Fitzpatrick, R. (2003). Distant past exercise in women: measures may be reliable, but are they valid? *Medicine & Science in Sports & Exercise*. 35 (5): 862-866.

De Bock, G.H., Kaptein, A.A., Touw-Otten, F. & Mulder, J.D. (1995). Health-related quality of life in patients with osteoarthritis in a family practice setting. *Arthritis Care and Research: The Official Journal of the Arthritis Health Professions Association.* 8 (2): 88-93.

De Boer, M.R., de Vet, H.C., Terwee, C.B., Moll, A.C., Völker-Dieben, H.J. & van Rens, G.H. (2005). Changes to the subscales of two vision-related quality of life questionnaires are proposed. *Journal of Clinical Epidemiology*. 58 (12): 1260-1268.

De Carvalho, L.H.J., Temponi, E.F., Soares, L.F.M., Goncalves, M.B.J. & Costa, L.P. (2014). Physical activity after distal femur osteotomy for the treatment of lateral compartment knee osteoarthritis. *Knee Surgery, Sports Traumatology, Arthroscopy.* 22 (7): 1607-1611.

De Groot, I., Bussmann, J., Stam, H. & Verhaar, J. (2008). Actual everyday physical activity in patients with end-stage hip or knee osteoarthritis compared with healthy controls. *Osteoarthritis and Cartilage.* 16 (4): 436-442.

De Hollander, E.L., Zwart, L., De, S.I. & Wendel-Vos, W. (2012). The SQUASH was a more valid tool than the OBiN for categorizing adults according to the Dutch physical activity and the combined guideline. *Journal of clinical epidemiology*. 65 (1): 73-81.

De Jonge, L., Nguyen, T., Smith, S.R., Zachwieja, J.J., Roy, H.J. & Bray, G.A. (2001). Prediction of energy expenditure in a whole body indirect calorimeter at both low and high levels of physical activity. *International Journal of Obesity & Related Metabolic Disorders.* 25 (7): 929.

De Vet, H.C., Terwee, C.B., Mokkink, L.B., Knol, D.L. (2011). *Measurement in medicine.* Practical guides to biostatistics and epidemiology Cambridge, United Kingdom: Cambridge University Press.

De Vet, H.C., Terwee, C.B. & Bouter, L.M. (2003). Clinimetrics and psychometrics: two sides of the same coin. *Journal of clinical epidemiology.* 56 (12): 1146-1147.

De Vet, H.C., Terwee, C.B., Knol, D.L. & Bouter, L.M. (2006). When to use agreement versus reliability measures. *Journal of clinical epidemiology.* 59 (10): 1033-1039.

De Vellis, R.F. (2006). Classical test theory. *Medical care.* 44 (11 Suppl 3): S50-9.

Delbaere, K., Hauer, K. & Lord, S.R. (2010). Evaluation of the incidental and planned activity questionnaire for older people. *British journal of sports medicine*. 44 (14): 1029-34.

Delucchi, K.L. & Bostrom, A. (2004). Methods for analysis of skewed data distributions in psychiatric clinical studies: working with many zero values. *American Journal of Psychiatry.* 161 (7): 1159-1168.

Deng, H.B., Macfarlane, D.J., Thomas, G.N., Lao, X.Q., Jiang, C.Q., Cheng, K.K. & Lam, T.H. (2008). Reliability and validity of the IPAQ-Chinese: the Guangzhou Biobank Cohort study. *Medicine & Science in Sports & Exercise.* 40 (2): 303-7.

Department of British Communities and Local Government (2012) *English indices of deprivation*. Department of British Communities and Local Government, London, England.

Department of Health (DH). (2004) *At Least Five a Week. Evidence on the Impact of Physical Activity and Its Relationship to Health. A report from the Chief Medical Officer*. Department of Health, London, UK.

Deslandes, A., Moraes, H., Ferreira, C., Veiga, H., Silveira, H., Mouta, R., Pompeu, F.A., Coutinho, E.S. & Laks, J. (2009). Exercise and mental health: many reasons to move. *Neuropsychobiology.* 59 (4): 191-198.

Dieppe, P.A., Cushnaghan, J. & Shepstone, L. (1997). The Bristol 'OA500'study: progression of osteoarthritis (OA) over 3 years and the relationship between clinical and radiographic changes at the knee joint. *Osteoarthritis and Cartilage.* 5 (2): 87-97.

Dieppe, P., Cushnaghan, J., Tucker, M., Browning, S. & Shepstone, L. (2000). The Bristol 'OA500 study': progression and impact of the disease after 8 years. *Osteoarthritis and Cartilage.* 8 (2): 63-68.

Dieppe, P.A. & Lohmander, L.S. (2005). Pathogenesis and management of pain in osteoarthritis. *The Lancet.* 365 (9463): 965-973.

DiLorenzo, T.M., Bargman, E.P., Stucky-Ropp, R., Brassington, G.S., Frensch, P.A. & LaFontaine, T. (1999). Long-term effects of aerobic exercise on psychological outcomes. *Preventive medicine.* 28 (1): 75-85.

Dinger, M.K., Oman, F., Taylor, E.L., Vesely, S.K. & Able, J. (2004). Stability and convergent validity of the Physical Activity Scale for the Elderly (PASE). *Journal of Sports Medicine and Physical Fitness*. 44 (2): 186-192.

Doble, S.E. & Fisher, A.G. (1998). The dimensionality and validity of the Older Americans Resources and Services (OARS) Activities of Daily Living (ADL) Scale. *Journal of outcome measurement.* 2 (1): 4-24.

Dobson, F., Hinman, R., Hall, M., Terwee, C., Roos, E.M. & Bennell, K. (2012). Measurement properties of performance-based measures to assess physical function in hip and knee osteoarthritis: a systematic review. *Osteoarthritis and Cartilage.* 20 (12): 1548-1562.

Donald, I.P. & Foy, C. (2004). A longitudinal study of joint pain in older people. *Rheumatology.* 43 (10): 1256-1260.

Dunlop, D.D., Song, J., Semanik, P.A., Chang, R.W., Sharma, L., Bathon, J.M., Eaton, C.B., Hochberg, M.C., Jackson, R.D. & Kwoh, C.K. (2011a). Objective physical activity measurement in the osteoarthritis initiative: Are guidelines being met? *Arthritis & Rheumatism.* 63 (11): 3372-3382.

Dunlop, D.D., Song, J., Semanik, P.A., Sharma, L. & Chang, R.W. (2011). Physical activity levels and functional performance in the osteoarthritis initiative: a graded relationship. *Arthritis & Rheumatism.* 63 (1): 127-136.

Dunn, A.L., Trivedi, M.H., Kampert, J.B., Clark, C.G. & Chambliss, H.O. (2005). Exercise treatment for depression: efficacy and dose response. *American Journal of Preventive Medicine.* 28 (1): 1-8.

Dyrstad, S.M., Hansen, B.H., Holme, I.M. & Anderssen, S.A. (2014). Comparison of self-reported versus accelerometer-measured physical activity. *Medicine & Science in Sports & Exercise.* 46 (1): 99-106.

Dziedzic, K.S., Healey, E.L., Porcheret, M., Ong, B., Main, C.J., Jordan, K.P., Lewis, M., Edwards, J.J., Jinks, C., Morden, A., McHugh, G.A., Ryan, S., Finney, A., Jowett, S., Oppong, R., Afolabi, E., Pushpa-Rajah, A., Handy, J., Clarkson, K., Mason, E., Whitehurst, T., Hughes, R.W., Croft, P.R. & Hay, E.M. (2014). Implementing the NICE osteoarthritis guidelines: a mixed methods study and cluster randomised trial of a model osteoarthritis consultation in primary care - the Management of OsteoArthritis In Consultations (MOSAICS) study protocol. *Implementation science: IS.* 9 (1): 95.

Dziedzic, K.S., Hill, S., Nicholls, E., Hammond, A., Myers, H., Whitehurst, T., Bailey, J., Clements, C., Whitehurst, D.G., Jowett, S., Handy, J., Hughes, R.W., Thomas, E. & Hay, E.M. (2011). Self management, joint protection and exercises

in hand osteoarthritis: a randomised controlled trial with cost effectiveness analyses. *BMC musculoskeletal disorders.* 12: 156-2474-12-156.

Edwards, P., Roberts, I., Clarke, M., DiGuiseppi, C., Pratap, S., Wentz, R. & Kwan, I. (2002). Increasing response rates to postal questionnaires: systematic review. *BMJ* (Clinical research ed.). 324 (7347): 1183.

Effing, T., Zielhuis, G., Kerstjens, H., van der Valk, P. & van der Palen, J. (2011). Community based physiotherapeutic exercise in COPD self-management: a randomised controlled trial. *Respiratory medicine.* 105 (3): 418-426.

Ettinger, W.H., Burns, R., Messier, S.P., Applegate, W., Rejeski, W.J., Morgan, T., Shumaker, S., Berry, M.J., O'Toole, M. & Monu, J. (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 (1): 25-31.

Ettinger, W.H., Davis, M.A., Neuhaus, J.M. & Mallon, K.P. (1994). Long-term physical functioning in persons with knee osteoarthritis from NHANES I: effects of comorbid medical conditions. *Journal of clinical epidemiology.* 47 (7): 809-815.

Ewald, B., Attia, J. & McElduff, P. (2013). How Many Steps are Enough? Dose Response Curves for Pedometer Steps and Multiple Health Markers in a Community Based Sample of Older Australians. *Journal of physical activity & health.*

Faber, M.J., Bosscher, R.J. & van Wieringen, P.C. (2006). Clinimetric properties of the performance-oriented mobility assessment. *Physical Therapy.* 86 (7): 944-954.

Fava, G.A. & Belaise, C. (2005). A discussion on the role of clinimetrics and the misleading effects of psychometric theory. *Journal of clinical epidemiology.* 58 (8): 753-756.

Felson, D.T. & Chaisson, C.E. (1997). 2 Understanding the relationship between body weight and osteoarthritis. *Bailliere's Clinical Rheumatology.* 11 (4): 671-681.

Felson, D.T., Chaisson, C.E., Hill, C.L., Totterman, S.M., Gale, M.E., Skinner, K.M., Kazis, L. & Gale, D.R. (2001). The association of bone marrow lesions with pain in knee osteoarthritis. *Annals of Internal Medicine.* 134 (7): 541-549.

Felson, D.T., Lawrence, R.C., Dieppe, P.A., Hirsch, R., Helmick, C.G., Jordan, J.M., Kington, R.S., Lane, N.E., Nevitt, M.C., Zhang, Y., Sowers, M., McAlindon, T., Spector, T.D., Poole, A.R., Yanovski, S.Z., Ateshian, G., Sharma, L., Buckwalter, J.A., Brandt, K.D. & Fries, J.F. (2000). Osteoarthritis: new insights. Part 1: the disease and its risk factors. *Annals of Internal Medicine.* 133 (8): 635-646.

Felson, D.T., Naimark, A., Anderson, J., Kazis, L., Castelli, W. & Meenan, R.F. (1987). The prevalence of knee osteoarthritis in the elderly. The Framingham Osteoarthritis Study.*Arthritis & Rheumatism.* 30 (8): 914-918.

Felson, D.T., Niu, J., Clancy, M., Sack, B., Aliabadi, P. & Zhang, Y. (2007). Effect of recreational physical activities on the development of knee osteoarthritis in older adults of different weights: the Framingham Study. *Arthritis Care & Research.* 57 (1): 6-12.

Felson, D.T., Niu, J., Yang, T., Torner, J., Lewis, C.E., Aliabadi, P., Sack, B., Sharma, L., Guermazi, A. & Goggins, J. (2013). Physical activity, alignment and knee osteoarthritis: data from MOST and the OAI. *Osteoarthritis and Cartilage.*

Fentem, P.H. (1994). ABC of sports medicine. Benefits of exercise in health and disease. *BMJ (Clinical research ed.).* 308 (6939): 1291-1295.

Fernandes, L., Hagen, K.B., Bijlsma, J.W., Andreassen, O., Christensen, P., Conaghan, P.G., DHerty, M., Geenen, R., Hammond, A., Kjeken, I., Lohmander, L.S., Lund, H., Mallen, C.D., Nava, T., Oliver, S., Pavelka, K., Pitsillidou, I., da Silva, J.A., de la Torre, J., Zanoli, G., Vliet Vlieland, T.P. & European League Against Rheumatism (EULAR) (2013). EULAR recommendations for the non-pharmacological core management of hip and knee osteoarthritis. *Annals of the Rheumatic Diseases.* 72 (7): 1125-1135.

Fjeldsoe, B.S., Winkler, E.A.H., Marshall, A.L., Eakin, E.G. & Reeves, M.M. (2013). Active adults recall their physical activity differently to less active adults: test-retest reliability and validity of a physical activity survey. *Health Promotion Journal of Australia.* 24 (1): 26-31.

Field, A. (2012). *Discovering statistics using SPSS.* Sage Publications Limited; Boston, United States.

Finney, A., Porcheret, M., Grime, J., Jordan, K.P., Handy, J., Healey, E., Ryan, S., Jester, R. & Dziedzic, K. (2013). Defining the Content of an Opportunistic Osteoarthritis Consultation With Primary Health Care Professionals: A Delphi Consensus Study. *Arthritis care & research.* 65 (6): 962-968.

Finney, A., Healey, E., Lewis, M., Ryan, S. & Dziedzic, K. (2013). Multisite peripheral joint pain in the community: prevalence and impact in the North West midlands, United Kingdom (UK). A*nnals of the rheumatic diseases.* 72: 1042-1043.

Fisher, N.E., Killampalli, V.V., Kundra, R.K., Jagodzinski, N.A., Mathur, K. & Reading, A.D. (2011). Sporting and physical activity following hip resurfacing. *International orthopaedics.* 35 (7): 977-980.

Fleiss, J.L. & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and psychological measurement*. 33(3), 613-619.

Fletcher, R.H., Fletcher, G.S. & Fletcher, S.W. (2013). *Clinical epidemiology: the essentials.* 5th edn. London, United Kingdom: Lippincott Williams & Wilkins.

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): 1162-1167.

Ford, E.S., Kohl, H.W., Mokdad, A.H. & Ajani, U.A. (2005). Sedentary behavior, physical activity, and the metabolic syndrome among US adults. *Obesity research.* 13 (3): 608-614.

Food and Drug Administration (2009). Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. *Fed Regist.* 74 (235): 65132-65133.

Ford, E.S., Merritt, R.K., Heath, G.W., Powell, K.E., Washburn, R.A., Kriska, A. & Haile, G. (1991). Physical activity behaviors in lower and higher socioeconomic status populations. *American Journal of Epidemiology.* 133 (12): 1246-1256.

Forsen, L., Loland, N.W., Vuillemin, A., Chinapaw, M.J., van Poppel, M.N., Mokkink, L.B., van Mechelen, W. & Terwee, C.B. (2010). Self-administered

physical activity questionnaires for the elderly: a systematic review of measurement properties. *Sports medicine.* 40 (7): 601-623.

Foster, N.E., Healey, E.L., Holden, M.A., Nicholls, E., Whitehurst, D.G., Jowett, S., Jinks, C., Roddy, E., Hay, E.M. & BEEP trial team (2014). A multicentre, pragmatic, parallel group, randomised controlled trial to compare the clinical and cost-effectiveness of three physiotherapy-led exercise intervention ns for knee osteoarthritis in older adults: the BEEP trial protocol (ISRCTN: 93634563). *BMC Musculoskeletal Disorders.* 15:254.

Fowler, F.J. (2014). *Survey research methods.* vol. 1. Boston, United States; Sage publications.

Fransen, M. & Edmonds, J. (1999). Reliability and validity of the EuroQol in patients with osteoarthritis of the knee. *Rheumatology (Oxford, England).* 38 (9): 807-813.

Fransen, M. & McConnell, S. (2008). Exercise for osteoarthritis of the knee. *The Cochrane database of systematic reviews.* (4):CD004376.

Fransen, M., McConnell, S., Hernandez-Molina, G. & Reichenbach, S. (2014). Exercise for osteoarthritis of the hip. *The Cochrane database of systematic reviews.* 4:CD007912.

Fransen, M., Nairn, L., Winstanley, J., Lam, P. & Edmonds, J. (2007). Physical activity for osteoarthritis management: a randomized controlled clinical trial evaluating hydrotherapy or Tai Chi classes. *Arthritis Care & Research.* 57 (3): 407-414.

Fransen, M., Su, S., Harmer, A., Blyth, F.M., Naganathan, V., Sambrook, P., Le Couteur, D. & Cumming, R.G. (2014). A longitudinal study of knee pain in older men: Concord Health and Ageing in Men Project. *Age & Ageing.* 43 (2): 206-212.

Freene, N., Waddington, G., Chesworth, W., Davey, R. & Cochrane, T. (2014). Validating two self-report physical activity measures in middle-aged adults completing a group exercise or home-based physical activity program. Journal of Science & Medicine in Sport. 17 (6): 611-616.

Friedenreich, C.M., Neilson, H.K. & Lynch, B.M. (2010). State of the epidemiological evidence on physical activity and cancer prevention. *European journal of cancer.* 46 (14): 2593-2604.

Fries, J., Bruce, B. & Cella, D. (2005). The promise of PROMIS: using item response theory to improve assessment of patient-reported outcomes. *Clinical and experimental rheumatology.* 23 (5): S53.

Galvao, D.A. & Newton, R.U. (2005). Review of exercise intervention studies in cancer patients. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology.* 23 (4): 899-909.

Gandek, B., Ware, J.E., Aaronson, N.K., Apolone, G., Bjorner, J.B., Brazier, J.E., Bullinger, M., Kaasa, S., Leplege, A. & Prieto, L. (1998). Cross-validation of item selection and scoring for the SF-12 Health Survey in nine countries: results from the IQOLA Project. *Journal of clinical epidemiology.* 51 (11): 1171-1178.

Gandhi, S.K., Warren Salmon, J., Zhao, S.Z., Lambert, B.L., Gore, P.R. & Conrad, K. (2001). Psychometric evaluation of the 12-item short-form health survey (SF-12) in osteoarthritis and rheumatoid arthritis clinical trials. *Clinical therapeutics.* 23 (7): 1080-1098.

Ganz, R., Parvizi, J., Beck, M., Leunig, M., Notzli, H. & Siebenrock, K.A. (2003). Femoroacetabular impingement: a cause for osteoarthritis of the hip. *Clinical orthopaedics and related research.* (417) (417): 112-120.

Garber, C.E., Blissmer, B., Deschenes, M.R., Franklin, B.A., Lamonte, M.J., Lee, I.M., Nieman, D.C., Swain, D.P. & American College of Sports Medicine (2011). American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. *Medicine and science in sports and exercise.* 43 (7): 1334-1359.

Garcia-Aymerich, J., Lange, P., Benet, M., Schnohr, P. & Antó, J.M. (2006). Regular physical activity reduces hospital admission and mortality in chronic obstructive pulmonary disease: a population based cohort study. *Thorax.* 61 (9): 772-778.

Gargon, E., Williamson, P.R., Altman, D.G., Blazeby, J.M. & Clarke, M. (2014). The COMET Initiative database: progress and activities from 2011 to 2013. *Trials.* 15 (1): 279.

Gelber, A.C., Hochberg, M.C., Mead, L.A., Wang, N.Y., Wigley, F.M. & Klag, M.J. (2000). Joint injury in young adults and risk for subsequent knee and hip osteoarthritis. *Annals of Internal Medicine.* 133 (5): 321-328.

Gerber, Y., Myers, V., Goldbourt, U., Benyamini, Y., Scheinowitz, M. & Drory, Y. (2011). Long-term trajectory of leisure time physical activity and survival after first myocardial infarction: a population-based cohort study. *European journal of epidemiology.* 26 (2): 109-116.

Giacomantonio, N.B., Bredin, S.S., Foulds, H.J. & Warburton, D.E. (2013). A systematic review of the health benefits of exercise rehabilitation in persons living with atrial fibrillation. *Canadian Journal of Cardiology.* 29 (4): 483-491.

Giannuzzi, P., Saner, H., Bjornstad, H., Fioretti, P., Mendes, M., Cohen-Solal, A., Dugmore, L., Hambrecht, R., Hellemans, I., McGee, H., Perk, J., Vanhees, L., Veress, G. & Working Group on Cardiac Rehabilitation and Exercise Physiology of the European Society of Cardiology (2003). Secondary prevention through cardiac rehabilitation: position paper of the Working Group on Cardiac Rehabilitation and Exercise Physiology of the European Society of Cardiology. *European heart journal.* 24 (13): 1273-1278.

Gilchrist, F., Rodd, H., Deery, C. & Marshman, Z. (2014). Assessment of the quality of measures of child oral health-related quality of life. *BMC oral health*. 14 : 40-6831-14-40.

Gillam, S. & Steel, N. (2013). The Quality and Outcomes Framework—where next? *BMJ: British Medical Journal.* 346.

Giraudeau, B. & Mary, J. (2001). Planning a reproducibility study: how many subjects and how many replicates per subject for an expected width of the 95 per cent confidence interval of the intraclass correlation coefficient. *Statistics in medicine.* 20 (21): 3205-3214.

Gleeson, M., Bishop, N.C., Stensel, D.J., Lindley, M.R., Mastana, S.S. & Nimmo, M.A. (2011). The anti-inflammatory effects of exercise: mechanisms and implications for the prevention and treatment of disease. *Nature Reviews Immunology.* 11 (9): 607-615.

Goodyear, L.J. & Kahn, B.B. (1998). Exercise, glucose transport, and insulin sensitivity. *Annual Review of Medicine.* 49: 235-261.

Greenland, S. & Robins, J.M. (1988). Conceptual problems in the definition and interpretation of attributable fractions. *American Journal of Epidemiology.* 128 (6): 1185-1197.

Grimes, D.A. & Schulz, K.F. (2002). Bias and causal associations in observational research. *The Lancet.* 359 (9302): 248-252.

Grimm, E.K., Swartz, A.M., Hart, T., Miller, N.E. & Strath, S.J. (2012). Comparison of the IPAQ-Short Form and accelerometry predictions of physical activity in older adults. *Journal of Ageing & Physical Activity*. 20 (1): 64-79.

Groenwold, R.H., Van Deursen, A.M., Hoes, A.W. & Hak, E. (2008). Poor quality of reporting confounding bias in observational intervention studies: a systematic review. *Annals of Epidemiology.* 18 (10): 746-751.

Groot, I.B., Bussmann, H.J., Stam, H.J. & Verhaar, J.A. (2008). Small increase of actual physical activity 6 months after total hip or knee arthroplasty. *Clinical orthopaedics and related research*. 466 (9): 2201-2208.

Guyatt, G., Walter, S. & Norman, G. (1987). Measuring change over time: assessing the usefulness of evaluative instruments. *Journal of chronic diseases*. 40 (2): 171-178.

Hadler, N.M., Gillings, D.B., Imbus, H.R., Levitin, P.M., Makuc, D., Utsinger, P.D., Yount, W.J., Slusser, D. & Moskovitz, N. (1978). Hand structure and function in an industrial setting. *Arthritis & Rheumatism.* 21 (2): 210-220.

Hagiwara, A., Ito, N., Sawai, K. & Kazuma, K. (2008). Validity and reliability of the Physical Activity Scale for the Elderly (PASE) in Japanese elderly people. *Geriatrics & gerontology international*. 8 (3): 143-51.

Halpin, D. (2004). NICE guidance for COPD. *Thorax.* 59 (3): 181-182.

Hammill, B.G., Curtis, L.H., Schulman, K.A. & Whellan, D.J. (2010). Relationship between cardiac rehabilitation and long-term risks of death and myocardial infarction among elderly Medicare beneficiaries. *Circulation.* 121 (1): 63-70.

Hamilton, M.T. & Booth, F.W. (2000). Skeletal muscle adaptation to exercise: a century of progress. *Journal of applied physiology.* 88 (1): 327-331.

Handschin, C. & Spiegelman, B.M. (2008). The role of exercise and PGC1α in inflammation and chronic disease. *Nature.* 454 (7203): 463-469.

Hannan, M.T., Felson, D.T. & Pincus, T. (2000). Analysis of the discordance between radiographic changes and knee pain in osteoarthritis of the knee. *The Journal of rheumatology.* 27 (6): 1513-1517.

Hansen, A.W., Dahl-Petersen, I., Helge, J.W., Brage, S., Gronbaek, M. & Flensborg-Madsen, T. (2014). Validation of an Internet-based long version of the International Physical Activity Questionnaire in Danish adults using combined accelerometry and heart rate monitoring. *Journal of Physical Activity & Health.* 11 (3): 654-664.

Harada, N.D., Chiu, V., King, A.C. & Stewart, A.L. (2001). An evaluation of three self-report physical activity instruments for older adults. *Medicine & Science in Sports & Exercise*. 33 (6): 962-70.

Harding, P., Holland, A.E., Delany, C. & Hinman, R.S. (2014). Do activity levels increase after total hip and knee arthroplasty? *Clinical Orthopaedics & Related Research.* 472 (5): 1502-1511.

Harris, T.J., Owen, C.G., Victor, C.R., Adams, R., Ekelund, U. & Cook, D.G. (2009). A comparison of questionnaire, accelerometer, and pedometer: measures in older people. *Medicine & Science in Sports & Exercise*. 41 (7): 1392-402.

Hart, T.L., Ainsworth, B.E. & Tudor-Locke, C. (2011). Objective and subjective measures of sedentary behavior and physical activity. *Med Sci Sports Exerc.* 43 (3): 449-456.

Harvey, S.B., Hotopf, M., Overland, S. & Mykletun, A. (2010). Physical activity and common mental disorders. *The British journal of psychiatry: the journal of mental science.* 197 (5): 357-364.

Haskell, W.L., Lee, I., Pate, R.R., Powell, K.E., Blair, S.N., Franklin, B.A., Macera, C.A., Heath, G.W., Thompson, P.D. & Bauman, A. (2007). Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. *Medicine and science in sports and exercise.* 39 (8): 1423.

Hassmén, P., Koivula, N. & Uutela, A. (2000). Physical exercise and psychological well-being: a population study in Finland. *Preventive medicine.* 30 (1): 17-25.

Hayden, J.A., Van Tulder, M.W. & Tomlinson, G. (2005). Systematic review: strategies for using exercise therapy to improve outcomes in chronic low back pain. *Annals of Internal Medicine.* 142 (9): 776-785.

Hayami, T., Pickarski, M., Zhuo, Y., Wesolowski, G.A., Rodan, G.A. & Duong, L.T. (2006). Characterization of articular cartilage and subchondral bone changes in the rat anterior cruciate ligament transection and meniscectomized models of osteoarthritis. *Bone.* 38 (2): 234-243.

Haywood, K.L., Garratt, A.M., Jordan, K.P., Healey, E.L. & Packham, J.C. (2010). Evaluation of ankylosing spondylitis quality of life (EASi-QoL): reliability and validity of a new patient-reported outcome measure. *The Journal of rheumatology.* 37 (10): 2100-2109.

Haywood, K.L., Wilson, R., Staniszewska, S. & Salek, S. (2016). Using PROMs in Healthcare: Who Should Be in the Driving Seat—Policy Makers, Health Professionals, Methodologists or Patients? *The patient: patient-centered outcomes research*. 9 (6): 495-498.

Healy, G.N., Dunstan, D.W., Salmon, J., Cerin, E., Shaw, J.E., Zimmet, P.Z. & Owen, N. (2007). Objectively measured light-intensity physical activity is independently associated with 2-h plasma glucose. *Diabetes care.* 30 (6): 1384-1389.

Healy, G.N., Dunstan, D.W., Salmon, J., Cerin, E., Shaw, J.E., Zimmet, P.Z. & Owen, N. (2008). Breaks in sedentary time: beneficial associations with metabolic risk. *Diabetes care.* 31 (4): 661-666.

Heesch, K.C., Hill, R.L., van, J.G., JG & Brown, W.J. (2011). Are Active Australia physical activity questions valid for older adults? *Journal of Science & Medicine in Sport*. 14 (3): 233-7.

Heesch, K.C., van Uffelen, J. & Brown, W.J. (2014). How do older adults respond to active Australia physical activity questions? Lessons from cognitive interviews. *Journal of Aging & Physical Activity.* 22 (1): 74-86.

Helmerhorst, H.J., Wijndaele, K., Brage, S., Wareham, N.J. & Ekelund, U. (2009). Objectively measured sedentary time may predict insulin resistance independent of moderate-and vigorous-intensity physical activity. *Diabetes.* 58 (8): 1776-1779.

Helmick, C.G., Felson, D.T., Lawrence, R.C., Gabriel, S., Hirsch, R., Kwoh, C.K., Liang, M.H., Kremers, H.M., Mayes, M.D. & Merkel, P.A. (2008). Estimates of the prevalence of arthritis and other rheumatic conditions in the United States: Part I. *Arthritis & Rheumatism.* 58 (1): 15-25.

Hennekens, C.H., Buring, J.E. & Mayrent, S.L. (1987). *Epidemiology in medicine.* Boston, United States: Little Brown.

Hermann, A., Ried-Larsen, M., Jensen, A.K., Holst, R., Andersen, L.B., Overgaard, S. & Holsgaard-Larsen, A. (2014). Low validity of the Sensewear Pro3 activity monitor compared to indirect calorimetry during simulated free living in patients with osteoarthritis of the hip. *BMC musculoskeletal disorders.* 15 (1): 43.

Hertogh, E.M., Monninkhof, E.M., Schouten, E.G., Peeters, P.H. & Schuit, A.J. (2008). Validity of the modified Baecke questionnaire: comparison with energy expenditure according to the doubly labeled water method. *International Journal of Behavioral Nutrition & Physical Activity*. 5.

Higgins, J.P. & Green, S. (2008). *Cochrane handbook for systematic reviews of interventions.* vol. 5. Wiley Online Library.

Hidding, A., van der Linden, S., Boers, M., Gielen, X., de Witte, L., Kester, A., Dijkmans, B. & Moolenburgh, D. (1993). Is group physical therapy superior to individualized therapy in ankylosing spondylitis? A randomized controlled trial. *Arthritis & Rheumatism.* 6 (3): 117-125.

Hochberg, M.C., Altman, R.D., April, K.T., Benkhalti, M., Guyatt, G., McGowan, J., Towheed, T., Welch, V., Wells, G. & Tugwell, P. (2012). American College of Rheumatology 2012 recommendations for the use of nonpharmacologic and

pharmacologic therapies in osteoarthritis of the hand, hip, and knee. *Arthritis care & research.* 64 (4): 465-474.

Holden, M.A., Nicholls, E.E., Young, J., Hay, E.M. & Foster, N.E. (2012). Role of exercise for knee pain: What do older adults in the community think? *Arthritis care & research.* 64 (10): 1554-1564.

Holden, M.A., Nicholls, E.E., Young, J., Hay, E.M. & Foster, N.E. (2009). UK-based physical therapists' attitudes and beliefs regarding exercise and knee osteoarthritis: Findings from a mixed-methods study. *Arthritis Care & Research.* 61 (11): 1511-1521.

Holloszy, J.O. & Coyle, E.F. (1984). Adaptations of skeletal muscle to endurance exercise and their metabolic consequences. *Journal of applied physiology: respiratory, environmental and exercise physiology.* 56 (4): 831-838.

Holm, B., Bandholm, T., Lunn, T.H., Husted, H., Aalund, P.K., Hansen, T.B. & Kehlet, H. (2014). Role of preoperative pain, muscle function, and activity level in discharge readiness after fast-track hip and knee arthroplasty. *Acta Orthopaedica*. 85 (5): 488-492.

Holsgaard-Larsen, A. & Roos, E. (2012). Objectively measured physical activity in patients with end stage knee or hip osteoarthritis. *Eur J Phys Rehabil Med.* 48 (4): 577-585.

Hoogeboom, T.J., den Broeder, A.A., de Bie, R.A. & van den Ende,Cornelia H M. (2013). Longitudinal impact of joint pain comorbidity on quality of life and activity levels in knee osteoarthritis: data from the Osteoarthritis Initiative. *Rheumatology.* 52 (3): 543-546.

Hootman, J.M. & Helmick, C.G. (2006). Projections of US prevalence of arthritis and associated activity limitations. *Arthritis & Rheumatism.* 54 (1): 226-229.

Hovis, K.K., Stehling, C., Souza, R.B., Haughom, B.D., Baum, T., Nevitt, M., McCulloch, C., Lynch, J.A. & Link, T.M. (2011). Physical activity is associated with magnetic resonance imageing-based knee cartilage T2 measurements in asymptomatic subjects with and those without osteoarthritis risk factors. *Arthritis & Rheumatism*. 63 (8): 2248-2256.

Hunter, D.J., Lo, G.H., Gale, D., Grainger, A.J., Guermazi, A. & Conaghan, P.G. (2008). The reliability of a new scoring system for knee osteoarthritis MRI and the validity of bone marrow lesion assessment: BLOKS (Boston Leeds Osteoarthritis Knee Score). *Annals of the Rheumatic Diseases.* 67 (2): 206-211.

Hunter, R.F., Tully, M.A., Donnelly, P., Stevenson, M. & Kee, F. (2014). Knowledge of UK physical activity guidelines: Implications for better targeted health promotion. *Preventive medicine.* 65: 33-39.

Hurley, M., Walsh, N., Mitchell, H., Nicholas, J. & Patel, A. (2012). Long-term outcomes and costs of an integrated rehabilitation program for chronic knee pain: A pragmatic, cluster randomized, controlled trial. *Arthritis care & research.* 64 (2): 238-247.

Hurtig-Wennlof, A., Hagstromer, M. & Olsson, L.A. (2010). The International Physical Activity Questionnaire modified for the elderly: aspects of validity and feasibility. *Public health nutrition*. 13 (11): 1847-54.

Iwamoto, J., Sato, Y., Takeda, T. & Matsumoto, H. (2011). Effectiveness of exercise for osteoarthritis of the knee: A review of the literature. *World journal of orthopedics.* 2 (5): 37-42.

Intema, F., Hazewinkel, H., Gouwens, D., Bijlsma, J., Weinans, H., Lafeber, F. & Mastbergen, S. (2010). In early OA, thinning of the subchondral plate is directly related to cartilage damage: results from a canine ACLT-meniscectomy model. *Osteoarthritis and Cartilage.* 18 (5): 691-698.

Isomaa, B., Almgren, P., Tuomi, T., Forsen, B., Lahti, K., Nissen, M., Taskinen, M.R. & Groop, L. (2001). Cardiovascular morbidity and mortality associated with the metabolic syndrome. *Diabetes care.* 24 (4): 683-689.

Jagger, C., Matthews, R., Spiers, N., Brayne, C., Comas-Herrera, A., Robinson, T., Lindesay, J. & Croft, P. (2006). *Compression or expansion of disability?: forecasting future disability levels under changing patterns of diseases. Wanless Social Care Review Research Report*. Leicester Nuffield Research Unit, University of Leicester, United Kingdom.

Jensen, C., Roos, E.M., Kjaersgaard-Andersen, P. & Overgaard, S. (2013). The effect of education and supervised exercise vs. education alone on the time to total hip replacement in patients with severe hip osteoarthritis. A randomized clinical trial protocol. *BMC Musculoskeletal Disorders.* 14 : 21.

Jeon, C.Y., Lokken, R.P., Hu, F.B. & Van Dam, R.M. (2007). Physical Activity of Moderate Intensity and Risk of Type 2 Diabetes A systematic review. *Diabetes care.* 30 (3): 744-752.

Jinks, C., Jordan, K. & Croft, P. (2003). Evaluation of a computer-assisted data entry procedure (including Teleform) for large-scale mailed surveys. *Computers in biology and medicine.* 33 (5): 425-437.

Jinks, C., Carter, P., Rhodes, C., Beech, R., Dziedzic, K., Hughes, R., Blackburn, S. & Ong, B.N. (2013). Sustaining patient and public involvement in research: a case study of a research centre. *Journal of care services management*. 7 (4): 146-154.

Johannsen, D.L., Calabro, M.A., Stewart, J., Franke, W., Rood, J.C. & Welk, G.J. (2010). Accuracy of armband monitors for measuring daily energy expenditure in healthy adults. *Medicine and Science in Sports and Exercise.* 42 (11): 2134-2140.

Johnson, T.P. & Wislar, J.S. (2012). Response rates and nonresponse errors in surveys. *JAMA: The Journal of the American Medical Association.* 307 (17): 1805-1806.

Jones, D.L., Bhanegaonkar, A.J., Billings, A.A., Kriska, A.M., Irrgang, J.J., Crossett, L.S. & Kwoh, C.K. (2012). Differences between actual and expected leisure activities after total knee arthroplasty for osteoarthritis. *The Journal of arthroplasty*. 27 (7): 1289-1296.

Jordan, K.P., Clarke, A.M., Symmons, D.P., Fleming, D., Porcheret, M., Kadam, U.T. & 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, K.P., Joud, A., Bergknut, C., Croft, P., Edwards, J.J., Peat, G., Petersson, I.F., Turkiewicz, A., Wilkie, R. & Englund, M. (2014). International comparisons of the consultation prevalence of musculoskeletal conditions using population-based healthcare data from England and Sweden. *Annals of the Rheumatic Diseases.* 73 (1): 212-218.

Jordan, K.P., Kadam, U.T., Hayward, R., Porcheret, M., Young, C. & Croft, P. (2010). Annual consultation prevalence of regional musculoskeletal problems in primary care: an observational study. *BMC Musculoskeletal Disorders.* 11 (1): 144.

Juhl, C., Christensen, R., Roos, E.M., Zhang, W. & Lund, H. (2014). Impact of Exercise Type and Dose on Pain and Disability in Knee Osteoarthritis: A Systematic Review and Meta-Regression Analysis of Randomized Controlled Trials. *Arthritis & rheumatology.* 66 (3): 622-636.

Kaipiainen-Seppanen, O., Aho, K. & Heliovaara, M. (1997). Incidence and prevalence of ankylosing spondylitis in Finland. *The Journal of rheumatology.* 24 (3): 496-499.

Kamaleri, Y., Natvig, B., Ihlebaek, C.M. & Bruusgaard, D. (2008). Localized or widespread musculoskeletal pain: does it matter? *Pain.* 138 (1): 41-46.

Kaminsky, L.A. (2006). *ACSM's resource manual for Guidelines for exercise testing and prescription*. (6th Ed) United States: Lippincott Williams & Wilkins.

Kapstad, H., Hanestad, B.R., Langeland, N., Rustoen, T. & Stavem, K. (2008). Cutpoints for mild, moderate and severe pain in patients with osteoarthritis of the hip or knee ready for joint replacement surgery. *BMC musculoskeletal disorders.* 9 : 55-2474-9-55.

Karachalios, T., Sarangi, P.P. & Newman, J.H. (1994). Severe varus and valgus deformities treated by total knee arthroplasty. *The journal of bone and joint surgery. British volume.* 76 (6): 938-942.

Kazis, L.E., Anderson, J.J. & Meenan, R.F. (1989). Effect sizes for interpreting changes in health status. *Medical care*. 27 (3): S178-S189.

Keller, S., Bann, C.M., Dodd, S.L., Schein, J., Mendoza, T.R. & Cleeland, C.S. (2004). Validity of the brief pain inventory for use in documenting the outcomes of patients with noncancer pain. *The Clinical journal of pain.* 20 (5): 309-318.

Kidd, B.L. (2006). Osteoarthritis and joint pain. *Pain.* 123 (1): 6-9.

King, G.A., Torres, N., Potter, C., Brooks, T.J. & Coleman, K.J. (2004). Comparison of activity monitors to estimate energy cost of treadmill exercise. *Medicine and science in sports and exercise.* 36 (7): 1244-1251.

Kjeken, I., Smedslund, G., Moe, R.H., Slatkowsky-Christensen, B., Uhlig, T. & Hagen, K.B. (2011). Systematic review of design and effects of splints and exercise programs in hand osteoarthritis. *Arthritis care & research.* 63 (6): 834-848.

Knols, R., Aaronson, N.K., Uebelhart, D., Fransen, J. & Aufdemkampe, G. (2005). Physical exercise in cancer patients during and after medical treatment: a systematic review of randomized and controlled clinical trials. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology.* 23 (16): 3830-3842.

Kovar, P.A., Allegrante, J.P., MacKenzie, C.R., Peterson, M.G., Gutin, B. & Charlson, M.E. (1992). Supervised Fitness Walking in Patients with Osteoarthritis of the KneeA Randomized, Controlled Trial. *Annals of Internal Medicine.* 116 (7): 529-534.

Kujala, U.M., Kaprio, J. & Sarna, S. (1994). Osteoarthritis of weight bearing joints of lower limbs in former elite male athletes. *BMJ (Clinical research ed.).* 308 (6923): 231-234.

Kuo, C.F., Grainge, M.J., Mallen, C., Zhang, W. & Doherty, M. (2015). Rising burden of gout in the UK but continuing suboptimal management: a nationwide population study. *Annals of the Rheumatic Diseases.* 74 (4): 661-667.

Kwak, L., Hagstromer, M., Sjostrom, M., (2012). Can the IPAQ-long be used to assess occupational physical activity? *Journal of physical activity & health*. 9: 1130-1137.

Kwakkel, G., van Peppen, R., Wagenaar, R.C., Wood Dauphinee, S., Richards, C., Ashburn, A., Miller, K., Lincoln, N., Partridge, C., Wellwood, I. & Langhorne, P. (2004). Effects of augmented exercise therapy time after stroke: a meta-analysis. *Stroke; a journal of cerebral circulation.* 35 (11): 2529-2539.

Ince, G., Sarpel, T., Durgun, B. & Erdogan, S. (2006). Effects of a multimodal exercise program for people with ankylosing spondylitis. *Physical Therapy.* 86 (7): 924-935.

Laforge, R.G., Rossi, J.S., Prochaska, J.O., Velicer, W.F., Levesque, D.A. & McHorney, C.A. (1999). Stage of regular exercise and health-related quality of life. *Preventive medicine.* 28 (4): 349-360.

Lam, C.L., Gandek, B., Ren, X.S. & Chan, M. (1998). Tests of scaling assumptions and construct validity of the Chinese (HK) version of the SF-36 Health Survey. Journal of clinical epidemiology. 51 (11): 1139-1147.

Lamonte, M.J. & Ainsworth, B.E. (2001). Quantifying energy expenditure and physical activity in the context of dose response. *Medicine and science in sports and exercise.* 33 (6; SUPP): S370-S378.

Landis, J.R. &Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics.* : 159-174.

Lane, N., Brandt, K., Hawker, G., Peeva, E., Schreyer, E., Tsuji, W. & Hochberg, M. (2011). OARSI-FDA initiative: defining the disease state of osteoarthritis. *Osteoarthritis and Cartilage.* 19 (5): 478-482.

Lanyon, P., Muir, K., DHerty, S. & DHerty, M. (2000). Assessment of a genetic contribution to osteoarthritis of the hip: sibling study. *BMJ (Clinical research ed.).* 321 (7270): 1179-1183.

Laurin, D., Verreault, R., Lindsay, J., MacPherson, K. & Rockwood, K. (2001). Physical activity and risk of cognitive impairment and dementia in elderly persons. *Archives of Neurology.* 58 (3): 498-504.

Lautenschlager, N.T., Cox, K.L., Flicker, L., Foster, J.K., van Bockxmeer, F.M., Xiao, J., Greenop, K.R. & Almeida, O.P. (2008). Effect of physical activity on cognitive function in older adults at risk for Alzheimer disease: a randomized trial. *Jama.* 300 (9): 1027-1037.

Lawrence, R.C., Felson, D.T., Helmick, C.G., Arnold, L.M., Choi, H., Deyo, R.A., Gabriel, S., Hirsch, R., Hochberg, M.C. & Hunder, G.G. (2008). Estimates of the prevalence of arthritis and other rheumatic conditions in the United States: Part II. *Arthritis & Rheumatism.* 58 (1): 26-35.

Lee, I.M. (2003). Physical activity and cancer prevention--data from epidemiologic studies. *Medicine and science in sports and exercise.* 35 (11): 1823-1827.

Lee, I., Shiroma, E.J., Lobelo, F., Puska, P., Blair, S.N. & Katzmarzyk, P.T. (2012). Effect of physical inactivity on major non-communicable diseases worldwide: an

analysis of burden of disease and life expectancy. *The Lancet.*380 (9838): 219-229.

Levinger, P., Menz, H.B., Wee, E., Feller, J.A., Bartlett, J.R. & Bergman, N.R. (2011). Physiological risk factors for falls in people with knee osteoarthritis before and early after knee replacement surgery. *Knee Surgery, Sports Traumatology, Arthroscopy*. 19 (7): 1082-1089.

Levine, J.A., Schleusner, S.J. & Jensen, M.D. (2000). Energy expenditure of nonexercise activity. *The American Journal of Clinical Nutrition.* 72 (6): 1451-1454.

Liang, M.H., Fossel, A.H. & Larson, M.G. (1990). Comparisons of five health status instruments for orthopedic evaluation. *Medical care*. 28 (7): 632-642.

Lim, H., Moon, Y. & Lee, M.S. (2005). Effects of home-based daily exercise therapy on joint mobility, daily activity, pain, and depression in patients with ankylosing spondylitis. *Rheumatology international.* 25 (3): 225-229.

Lin, W., Alizai, H., Joseph, G.B., Srikhum, W., Nevitt, M.C., Lynch, J.A., McCulloch, C.E. & Link, T.M. (2013). Physical activity in relation to knee cartilage T2 progression measured with 3 T MRI over a period of 4 years: data from the Osteoarthritis Initiative. *Osteoarthritis & Cartilage.* 21 (10): 1558-1566.

Lindholm, E., Brevinge, H., Bergh, C., Körner, U. & Lundholm, K. (2003). Relationships between self-reported health related quality of life and measures of standardized exercise capacity and metabolic efficiency in a middle-aged and aged healthy population. *Quality of Life Research.* 12 (5): 575-582.

Litwic, A., Edwards, M.H., Dennison, E.M. & Cooper, C. (2013). Epidemiology and burden of osteoarthritis. *British medical bulletin.* 105: 185-199.

Livesey, G. & Elia, M. (1988). Estimation of energy expenditure, net carbohydrate utilization, and net fat oxidation and synthesis by indirect calorimetry: evaluation of errors with special reference to the detailed composition of fuels. *The American Journal of Clinical Nutrition.* 47 (4): 608-628.

Livingstone, M., Prentice, A.M., Coward, W.A., Ceesay, S.M., Strain, J.J., McKenna, P.G., Nevin, G.B., Barker, M.E. & Hickey, R.J. (1990). Simultaneous measurement of free-living energy expenditure by the doubly labeled water method and heart-rate monitoring. *The American Journal of Clinical Nutrition.* 52 (1): 59-65.

Loeser, R.F. (2013). Ageing processes and the development of osteoarthritis. *Current opinion in rheumatology.* 25 (1): 108-113.

Lohmander, L.S., Englund, P.M., Dahl, L.L. & Roos, E.M. (2007). The long-term consequence of anterior cruciate ligament and meniscus injuries: osteoarthritis. *The American Journal of Sports Medicine.* 35 (10): 1756-1769.

Loureiro, A., Mills, P.M. & Barrett, R.S. (2013). Muscle weakness in hip osteoarthritis: A systematic review. *Arthritis Care & Research.* 65 (3): 340-352.

Lubbeke, A., Zimmermann-Sloutskis, D., Stern, R., Roussos, C., Bonvin, A., Perneger, T., Peter, R. & Hoffmeyer, P. (2014). Physical activity before and after primary total hip arthroplasty: a registry-based study*. Arthritis care & research*. 66 (2): 277-284.

Lytle, M.E., Vander Bilt, J., Pandav, R.S., Dodge, H.H. & Ganguli, M. (2004). Exercise level and cognitive decline: the MoVIES project. *Alzheimer Disease & Associated Disorders.* 18 (2): 57-64.

MacAuley, D. (1994). A history of physical activity, health and medicine. *Journal of the Royal Society of Medicine.* 87 (1): 32-35.

MacKinnon, D.P., Krull, J.L. & Lockwood, C.M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention Science.* 1 (4): 173-181.

Mader, U., Martin, B.W., Schutz, Y. & Marti, B. (2006). Validity of four short physical activity questionnaires in middle-aged persons. *Medicine & Science in Sports & Exercise*. 38 (7): 1255-66.

Malmberg, J., Miilunpalo, S., Pasanen, M., Vuori, I. & Oja, P. (2005). Characteristics of leisure time physical activity associated with risk of decline in perceived health—a 10-year follow-up of middle-aged and elderly men and women. *Preventive medicine.* 41 (1): 141-150.

Mansournia, M.A., Danaei, G., Forouzanfar, M.H., Mahmoodi, M., Jamali, M., Mansournia, N. & Mohammad, K. (2012). Effect of physical activity on functional performance and knee pain in patients with osteoarthritis: analysis with marginal structural models. *Epidemiology.* 23 (4): 631-640.

Martin, K.A., Rejeski, W.J., Miller, M.E., James, M.K., Ettinger, W.H.,Jr & Messier, S.P. (1999). Validation of the PASE in older adults with knee pain and physical disability. *Medicine & Science in Sports & Exercise.* 31 (5): 627-33.

Martin, K.R., Kuh, D., Harris, T.B., Guralnik, J.M., Coggon, D. & Wills, A.K. (2013). Body mass index, occupational activity, and leisure-time physical activity: an exploration of risk factors and modifiers for knee osteoarthritis in the 1946 British birth cohort. *BMC Musculoskeletal Disorders*. 14 : 219.

Martinsen, E.W. (2008). Physical activity in the prevention and treatment of anxiety and depression. *Nordic Journal of Psychiatry.* 62 (S47): 25-29.

Marx, R.G., Menezes, A., Horovitz, L., Jones, E.C. & Warren, R.F. (2003). A comparison of two time intervals for test-retest reliability of health status instruments. *Journal of clinical epidemiology.* 56 (8): 730-735.

Matthews, C.E., Ainsworth, B.E., Hanby, C., Pate, R.R., Addy, C., Freedson, P.S., Jones, D.A. & Macera, C.A. (2005). Development and testing of a short physical activity recall questionnaire. *Medicine and Science in Sports and Exercise.* 37 (6): 986-994.

McAlindon, T.E., Bannuru, R.R., Sullivan, M.C., Arden, N.K., Berenbaum, F., Bierma-Zeinstra, S.M., Hawker, G.A., Henrotin, Y., Hunter, D.J. & Kawaguchi, H. (2014). OARSI guidelines for the non-surgical management of knee osteoarthritis. *Osteoarthritis and Cartilage.* 22 (3): 363-388.

McAlindon, T.E., Cooper, C., Kirwan, J.R. & Dieppe, P.A. (1993). Determinants of disability in osteoarthritis of the knee. *Annals of the Rheumatic Diseases.* 52 (4): 258-262.

McArdle, W.D., Katch, F.I. & Katch, V.L. (2010). *Exercise physiology: Nutrition, energy, and human performance. USA:* Lippincott Williams & Wilkins.

McDonald, R.P. (1999). Test theory: A unified treatment. *UK:* Psychology Press.

McHorney, C.A. & Tarlov, A.R. (1995). Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Quality of life research: an international journal of quality of life aspects of treatment, care and rehabilitation.* 4 (4): 293-307.

Mead, G.E., Morley, W., Campbell, P., Greig, C.A., McMurdo, M. & Lawlor, D.A. (2009). Exercise for depression. *Cochrane Database Systematic Reviews.* 3.

Melillo, K.D., Futrell, M., Williamson, E., Chamberlain, C., Bourque, A.M., MacDonnell, M. & Phaneuf, J.P. (1996). Perceptions of physical fitness and exercise activity among older adults. *Journal of advanced nursing.* 23 (3): 542-547.

Merom, D., Delbaere, K., Cumming, R., Voukelatos, A., Rissel, C., Van Der Ploeg, H.P. & Lord, S.R. (2014). Incidental and Planned Exercise Questionnaire for seniors: validity and responsiveness. *Medicine & Science in Sports & Exercise.* 46 (5): 947-954.

Milanovic, Z., Pantelic, S., Trajkovic, N., Jorgic, B., Sporis, G. & Bratic, M. (2014). Reliability of the Serbian version of the International Physical Activity Questionnaire for older adults. *Clinical Interventions In Aging.* 9: 581-587.

Minor, M.A., Webel, R.R., Kay, D.R., Hewett, J.E. & Anderson, S.K. (1989). Efficacy of physical conditioning exercise in patients with rheumatoid arthritis and osteoarthritis. *Arthritis & Rheumatism.* 32 (11): 1396-1405.

Mock, V., Dow, K.H., Meares, C.J., Grimm, P.M., Dienemann, J.A., Haisfield-Wolfe, M.E., Quitasol, W., Mitchell, S., Chakravarthy, A. & Gage, I. (1997). Effects of exercise on fatigue, physical functioning, and emotional distress during radiation therapy for breast cancer. *Oncology nursing forum.* 24 (6): 991-1000.

Moher, D., Liberati, A., Tetzlaff, J. & Altman, D.G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of Internal Medicine.* 151 (4): 264-269.

Mokkink, L.B., Terwee, C.B., Patrick, D.L., Alonso, J., Stratford, P.W., Knol, D.L., Bouter, L.M. & de Vet, H.C. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of clinical epidemiology.* 63 (7): 737-745.

Montoye, H.J. (2000). Introduction: evaluation of some measurements of physical activity and energy expenditure. *Medicine and science in sports and exercise.* 32 (9; SUPP/1): S439-S441.

Moore, D.S., Ellis, R., Allen, P.D., Cherry, K.E., Monroe, P.A., O'Neil, C.E. & Wood, R.H. (2008). Construct validation of physical activity surveys in culturally diverse older adults: a comparison of four commonly used questionnaires. *Research Quarterly for Exercise & Sport*. 79 (1): 42-50.

Moreira-Silva, I., Santos, R., Abreu, S. & Mota, J. (2014). The effect of a physical activity program on decreasing physical disability indicated by musculoskeletal pain and related symptoms among workers: a pilot study*. International Journal of Occupational Safety & Ergonomics*. 20 (1): 55-64.Morris, F.L., Naughton, G.A., Gibbs, J.L., Carlson, J.S. & Wark, J.D. (1997). Prospective ten-month exercise intervention in premenarcheal girls: positive effects on bone and lean mass. *Journal of bone and mineral research.* 12 (9): 1453-1462.Morrow Jr, J.R., Krzewinski-Malone, J.A., Jackson, A.W., Bungum, T.J. & Fitzgerald, S.J. (2004). American adults' knowledge of exercise recommendations. *Research quarterly for exercise and sport.* 75 (3): 231-237.

Mozaffarian, D., Furberg, C.D., Psaty, B.M. & Siscovick, D. (2008). Physical activity and incidence of atrial fibrillation in older adults the cardiovascular health study. *Circulation.* 118 (8): 800-807.

Murphy, S.L. (2009). Review of physical activity measurement using accelerometers in older adults: considerations for research design and conduct. *Preventive medicine.* 48 (2): 108-114.

Murray, C.J., Richards, M.A., Newton, J.N., Fenton, K.A., Anderson, H.R., Atkinson, C., Bennett, D., Bernabé, E., Blencowe, H. & Bourne, R. (2013). UK health performance: findings of the Global Burden of Disease Study 2010. *The Lancet.* 381 (9871): 997-1020.

Naal, F.D., Impellizzeri, F.M. & Leunig, M. (2009). Which is the best activity rating scale for patients undergoing total joint arthroplasty? *Clinical orthopaedics and related research*. 467 (4): 958-965.

Naal, F.D., Impellizzeri, F.M., Loibl, M., Huber, M. & Rippstein, P.F. (2009a). Habitual physical activity and sports participation after total ankle arthroplasty. *The American Journal of Sports Medicine.* 37 (1): 95-102.

National Institute for Health and Care Excellence (2006). *Obesity: the prevention, identification, assessment and management of overweight and obesity in adults and children.* CG43edn. London: National Institute for Health and Care Excellence.

National Institute for Health and Care Excellence (2008). *Osteoarthritis: National Clinical Guideline For Care And Management In Adults.* CG59edn. London: National Institute for Health and Care Excellence.

National Institute for Health and Care Excellence (2008). *Occupational therapy and physical activity interventions to promote the mental wellbeing of older people*

in primary care and residential care. PH16edn. London: National Institute for Health and Care Excellence.

National Institute for Health and Care Excellence (2014). *Osteoarthritis. Care and management in adults.* CG177. London: National Institute for Health and Care Excellence.

Nelson, M.E., Rejeski, W.J., Blair, S.N., Duncan, P.W., Judge, J.O., King, A.C., Macera, C.A. & Castaneda-Sceppa, C. (2007). Physical activity and public health in older adults: recommendation from the American College of Sports Medicine and the American Heart Association. *Medicine and science in sports and exercise.* 39 (8): 1435.

Neogi, T. & Zhang, Y. (2013). Epidemiology of osteoarthritis. *Rheumatic diseases clinics of North America.* 39 (1): 1-19.

Nevitt, M.C. & Felson, D.T. (1996). Sex hormones and the risk of osteoarthritis in women: epidemiological evidence. *Annals of the Rheumatic Diseases.* 55 (9): 673-676.

Nevitt, M.C., Felson, D.T. & Lester, G. (2013). *The Osteoarthritis Initiative: The protocol for the cohort study.* University of California San Francisco: The Osteoarthritis Initiative.

Ngai, S.P.C., Cheung, R.T.H., Lam, P.L., Chiu, J.K.W. & Fung, E.Y.H. (2012). Validation and reliability of the Physical Activity Scale for the Elderly in Chinese population. *Journal of Rehabilitation Medicine.* 44 (5): 462-465.

Nulty, D.D. (2008). The adequacy of response rates to online and paper surveys: what can be done? *Assessment & Evaluation in Higher Education.* 33 (3): 301-314.

O'Reilly, S.C., Muir, K.R. & DHerty, M. (1999). Effectiveness of home exercise on pain and disability from osteoarthritis of the knee: a randomised controlled trial. *Annals of the Rheumatic Diseases.* 58 (1): 15-19.

Øiestad, B.E., Østerås, N., Frobell, R., Grotle, M., Brøgger, H. & Risberg, M.A. (2013). Efficacy of strength and aerobic exercise on patient-reported outcomes and structural changes in patients with knee osteoarthritis: study protocol for a randomized controlled trial. *BMC musculoskeletal disorders.* 14 (1): 266.

Oliveria, S.A., Felson, D.T., Cirillo, P.A., Reed, J.I. & Walker, A.M. (1999). Body weight, body mass index, and incident symptomatic osteoarthritis of the hand, hip, and knee. *Epidemiology.* 10 (2): 161-166.

Oliveria, S.A., Felson, D.T., Reed, J.I., Cirillo, P.A. & Walker, A.M. (1995). Incidence of symptomatic hand, hip, and knee osteoarthritis among patients in a health maintenance organization. *Arthritis & Rheumatism.* 38 (8): 1134-1141.

Ono, R., Hirata, S., Yamada, M., Nishiyama, T., Kurosaka, M. & Tamura, Y. (2007). Reliability and validity of the Baecke physical activity questionnaire in adult women with hip disorders. *BMC musculoskeletal disorders*. 8: 61.

Orgeta, V., Regan, C. & Orrell, M. (2010). Physical activity for improving cognition in older people with mild cognitive impairment. *Cochrane Database of Systematic Reviews 2010.* Issue 1.

Oxford Economics (2010). *The economic cost of arthritis for the UK Economy.* Final Report. Oxford, United Kingdom: Oxford Economics.

Packham, J.C., Jordan, K.P., Haywood, K.L., Garratt, A.M. & Healey, E.L. (2012). Evaluation of Ankylosing Spondylitis Quality of Life questionnaire: responsiveness of a new patient-reported outcome measure. *Rheumatology*. 51 (4): 707-714.

Pallant, J. (2011). SPSS survival manual: A step by step guide to data analysis using SPSS. Open University Press; Maidenhead, United Kingdom.

Partonen, T., Leppämäki, S., Hurme, J. & Lönnqvist, J. (1998). Randomized trial of physical exercise alone or combined with bright light on mood and health-related quality of life. *Psychological medicine.* 28 (06): 1359-1364.

Patrick, D.L., Burke, L.B., Gwaltney, C.J., Leidy, N.K., Martin, M.L., Molsen, E. & Ring, L. (2011). Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: part 2—assessing respondent understanding. *Value in Health.* 14 (8): 978-988.

Paul, A., Lewis, M., Shadforth, M.F., Croft, P.R., Van Der Windt, D.A. & Hay, E.M. (2004). A comparison of four shoulder-specific questionnaires in primary care. *Annals of the Rheumatic Diseases*. 63 (10): 1293-1299.

Pauwels, R.A., Buist, A.S., Calverley, P.M., Jenkins, C.R. & Hurd, S.S. (2012). Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. *American journal of respiratory and critical care medicine.* 163 (5).

Peat, G., McCarney, R. & Croft, P. (2001). Knee pain and osteoarthritis in older adults: a review of community burden and current use of primary health care. *Annals of the Rheumatic Diseases.* 60 (2): 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.

Pedersen, B.K. & Saltin, B. (2006). Evidence for prescribing exercise as therapy in chronic disease. *Scandinavian Journal of Medicine & Science in Sports.* 16 (S1): 3-63.

Pelletier, J., Martel-Pelletier, J. & Abramson, S.B. (2001). Osteoarthritis, an inflammatory disease: potential implication for the selection of new therapeutic targets. *Arthritis & Rheumatism.* 44 (6): 1237-1247.

Pelliccia, A., Maron, B.J., Di Paolo, F.M., Biffi, A., Quattrini, F.M., Pisicchio, C., Roselli, A., Caselli, S. & Culasso, F. (2005). Prevalence and clinical significance of left atrial remodeling in competitive athletes. *Journal of the American College of Cardiology.* 46 (4): 690-696.

Pereira, D., Peleteiro, B., Araujo, J., Branco, J., Santos, R. & Ramos, E. (2011). The effect of osteoarthritis definition on prevalence and incidence estimates: a systematic review.*Osteoarthritis and Cartilage.* 19 (11): 1270-1285.

Petersen, C.B., Grønbæk, M., Helge, J.W., Thygesen, L.C., Schnohr, P. & Tolstrup, J.S. (2012). Changes in physical activity in leisure time and the risk of myocardial infarction, ischemic heart disease, and all-cause mortality. *European journal of epidemiology.* 27 (2): 91-99.

Petrella, R. &Bartha, C. (2000). Home based exercise therapy for older patients with knee osteoarthritis: A randomized clinical trial. *Journal of Rheumatology*. 27 (9): 2215-2221.

Petruzzello, S.J., Landers, D.M., Hatfield, B.D., Kubitz, K.A. & Salazar, W. (1991). A meta-analysis on the anxiety-reducing effects of acute and chronic exercise. *Sports medicine.* 11 (3): 143-182.

Pettee, K., McClain, J.J., Lee, C.D., Swan, P.D., Alvar, B.A., Mitros, M.R. & Ainsworth, B.E. (2009). Evaluation of physical activity measures used in middle-aged women. *Medicine & Science in Sports & Exercise*. 41 (7): 1403-12.

Pham, T., van der Heijde, D., Altman, R., Anderson, J., Bellamy, N., Hochberg, M., Simon, L., Strand, V., Woodworth, T. & Dougados, M. (2004). OMERACT-OARSI initiative: Osteoarthritis Research Society International set of responder criteria for osteoarthritis clinical trials revisited. *Osteoarthritis and Cartilage.* 12 (5): 389-399.

Piepoli, M., Davos, C., Francis, D., Coats, A., Belardinelli, R., Purcaro, A., Dubach, P., Myers, J., Giannuzzi, P. & Temporelli, P. (2004). Exercise training meta-analysis of trials in patients with chronic heart failure (ExTraMATCH). *BMJ.* 328 (7433): 189-192.

Pisters, M. F., Veenhof, C., de Bakker, D. H., Schellevis, F. G., & Dekker, J. (2010). Behavioural graded activity results in better exercise adherence and more physical activity than usual care in people with osteoarthritis: a cluster-randomised trial. *Journal of Physiotherapy*, 56(1), 41-47.

Pitta, F., Troosters, T., Spruit, M.A., Probst, V.S., Decramer, M. & Gosselink, R. (2005). Characteristics of physical activities in daily life in chronic obstructive pulmonary disease. *American journal of respiratory and critical care medicine.* 171 (9): 972-977.

Plasqui, G. & Westerterp, K.R. (2007). Physical activity assessment with accelerometers: an evaluation against doubly labeled water. *Obesity.* 15 (10): 2371-2379.

Podsiadlo, D. & Richardson, S. (1991). The timed “Up & Go”: a test of basic functional mobility for frail elderly persons. *Journal of the American Geriatrics Society.* 39 (2): 142-148.

Pols, M.A., Peeters, P.H., Bueno-De-Mesquita, H.B., Ocke, M.C., Wentink, C.A., Kemper, H.C. & Collette, H.J. (1995). Validity and repeatability of a modified Baecke questionnaire on physical activity. *International journal of epidemiology*. 24 (2): 381-8.

Pols, M.A., Peeters, P.H., Kemper, H.C. & Collette, H.J. (1996). Repeatability and relative validity of two physical activity questionnaires in elderly women. *Medicine & Science in Sports & Exercise*. 28 (8): 1020-5.

Porcheret, M., Grime, J., Main, C. & Dziedzic, K. (2013). Developing a model osteoarthritis consultation: a Delphi consensus exercise. *BMC musculoskeletal disorders.* 14 : 25-2474-14-25.

Powell, L.E. & Myers, A.M. (1995). The Activities-specific Balance Confidence (ABC) Scale. *The journals of gerontology.Series A, Biological sciences and medical sciences.* 50A (1): M28-34.

Powell, K.E., Paluch, A.E. & Blair, S.N. (2011). Physical activity for health: What kind? How much? How intense? On top of what? *Public health.* 32 (1): 349.

Prince, S., Adamo, K., Hamel, M., Hardt, J., Gorber, S. & Tremblay, M. (2008). A comparison of direct versus self-report measures for assessing physical activity in

adults: a systematic review. *International Journal of Behavioral Nutrition and Physical Activity.* 5 (1): 56.

Proper, K.I., Singh, A.S., Van Mechelen, W. & Chinapaw, M.J. (2011). Sedentary behaviors and health outcomes among adults: a systematic review of prospective studies. *American Journal of Preventive Medicine.* 40 (2): 174-182.

Qu, X., Zhang, X., Zhai, Z., Li, H., Liu, X., Li, H., Liu, G., Zhu, Z., Hao, Y. & Dai, K. (2013). Association between physical activity and risk of fracture. *Journal of Bone and Mineral Research.*

Rabin, R. & Charro, F.d. (2001). EQ-SD: a measure of health status from the EuroQol Group. *Annals of Medicine.* 33 (5): 337-343.

Ramsey, D.K., Snyder-Mackler, L., Lewek, M., Newcomb, W. & Rudolph, K.S. (2007). Effect of anatomic realignment on muscle function during gait in patients with medial compartment knee osteoarthritis. *Arthritis Care & Research.* 57 (3): 389-397.

Reichenbach, S., Rutjes, A.W., Nuesch, E., Trelle, S. & Juni, P. (2010). Joint lavage for osteoarthritis of the knee. *Cochrane database of systematic reviews (Online*). (5) (5): CD007320.

Reijman, M., Pols, H., Bergink, A., Hazes, J., Belo, J., Lievense, A. & Bierma-Zeinstra, S. (2007). Body mass index associated with onset and progression of osteoarthritis of the knee but not of the hip: the Rotterdam Study. *Annals of the Rheumatic Diseases.* 66 (2): 158-162.

Rennie, K. & Jebb, S. (2005). Prevalence of obesity in Great Britain. *Obesity reviews.* 6 (1): 11-12.

Rennie, K.L. & Wareham, N.J. (1998). The validation of physical activity instruments for measuring energy expenditure: problems and pitfalls. *Public health nutrition.* 1 (4): 265-271.

Revicki, D. & Cella, D. (1997). Health status assessment for the twenty-first century: item response theory, item banking and computer adaptive testing. *Quality of Life Research.* 6 (6): 595-600.

Rezende, L.F., Rey-López, J.P., Matsudo, V.K. & do Luiz, O. (2014). Sedentary behavior and health outcomes among older adults: a systematic review. *BMC public health.* 14 (1): 333.

Ribera, A., Permanyer-Miralda, G., Alonso, J., Cascant, P., Soriano, N. & Brotons, C. (2006). Is psychometric scoring of the McNew Quality of Life after Myocardial Infarction questionnaire superior to the clinimetric scoring? A comparison of the two approaches. *Quality of life research.* 15 (3): 357-365.

Richardson, J.T. (2005). Instruments for obtaining student feedback: A review of the literature. *Assessment & Evaluation in Higher Education.* 30 (4): 387-415.

Richette, P., Corvol, M. & Bardin, T. (2003). Estrogens, cartilage, and osteoarthritis. *Joint Bone Spine.* 70 (4): 257-262.

Riise, T., Moen, B.E. & Nortvedt, M.W. (2003). Occupation, lifestyle factors and health-related quality of life: the Hordaland Health Study. *Journal of occupational and environmental medicine.* 45 (3): 324-332.

Rock, C.L., Doyle, C., Demark-Wahnefried, W., Meyerhardt, J., Courneya, K.S., Schwartz, A.L., Bandera, E.V., Hamilton, K.K., Grant, B. & McCullough, M. (2012). Nutrition and physical activity guidelines for cancer survivors. *CA: a cancer journal for clinicians.* 62 (4): 242-274.

Roddy, E., Zhang, W., DHerty, M., Arden, N.K., Barlow, J., Birrell, F., Carr, A., Chakravarty, K., Dickson, J. & Hay, E. (2005). Evidence-based recommendations for the role of exercise in the management of osteoarthritis of the hip or knee—the MOVE consensus. *Rheumatology.* 44 (1): 67-73.

Rognmo, Ø., Moholdt, T., Bakken, H., Hole, T., Mølstad, P., Myhr, N.E., Grimsmo, J. & Wisløff, U. (2012). Cardiovascular Risk of High-Versus Moderate-Intensity Aerobic Exercise in Coronary Heart Disease Patients Clinical Perspective. *Circulation.* 126 (12): 1436-1440.

Rolland, Y., Pillard, F., Klapouszczak, A., Reynish, E., Thomas, D., Andrieu, S., Rivière, D. & Vellas, B. (2007). Exercise Program for Nursing Home Residents with Alzheimer's Disease: A 1-Year Randomized, Controlled Trial. *Journal of the American Geriatrics Society.* 55 (2): 158-165.

Rosemann, T., Grol, R., Herman, K., Wensing, M. & Szecsenyi, J. (2008a). Association between obesity, quality of life, physical activity and health service utilization in primary care patients with osteoarthritis. *The international journal of behavioral nutrition and physical activity.* 5: 4-5868-5-4.

Rosemann, T., Joos, S., Laux, G., Gensichen, J. & Szecsenyi, J. (2007). Case management of arthritis patients in primary care: A cluster-randomized controlled trial. *Arthritis Care & Research.* 57 (8): 1390-1397.

Rosemann, T., Kuehlein, T., Laux, G. & Szecsenyi, J. (2007a). Osteoarthritis of the knee and hip: a comparison of factors associated with physical activity. *Clinical rheumatology.* 26 (11): 1811-1817.

Rosemann, T., Kuehlein, T., Laux, G. & Szecsenyi, J. (2008). Factors associated with physical activity of patients with osteoarthritis of the lower limb. *Journal of evaluation in clinical practice.* 14 (2): 288-293.

Rothman, K.J., Greenland, S. & Lash, T.L. (2008). *Modern epidemiology.* 3rd edn. Philadelphia, United States: Lippincott Williams & Wilkins.

Rothschild, B.M., Xiaoting, Z. & Martin, L.D. (2012). Osteoarthritis in the early avian radiation: Earliest recognition of the disease in birds. *Cretaceous Research.* 35:178-180.

Rovio, S., Kåreholt, I., Helkala, E., Viitanen, M., Winblad, B., Tuomilehto, J., Soininen, H., Nissinen, A. & Kivipelto, M. (2005). Leisure-time physical activity at midlife and the risk of dementia and Alzheimer's disease. *The Lancet Neurology.* 4 (11): 705-711.

Rubenstein, L.Z. (2006). Falls in older people: epidemiology, risk factors and strategies for prevention. *Age and Ageing.* 35 (suppl 2): ii37-ii41.

Rutjes, A.W., Nuesch, E., Sterchi, R., Kalichman, L., Hendriks, E., Osiri, M., Brosseau, L., Reichenbach, S. & Juni, P. (2009). Transcutaneous electrostimulation for osteoarthritis of the knee. *The Cochrane database of systematic reviews.* (4):CD002823.

Saether, R., Helbostad, J.L., Riphagen, I.I. & Vik, T. (2013). Clinical tools to assess balance in children and adults with cerebral palsy: a systematic review. Developmental Medicine & Child Neurology. 55 (11): 988-999.

Salaffi, F., Carotti, M., Stancati, A. & Grassi, W. (2005). Health-related quality of life in older adults with symptomatic hip and knee osteoarthritis: a comparison with

matched healthy controls. *Ageing clinical and experimental research.* 17 (4): 255-263.

Sallis, J.F. & Saelens, B.E. (2000). Assessment of physical activity by self-report: status, limitations, and future directions. *Research quarterly for exercise and sport.* 71 (2 Suppl): S1-14.

Santos-Magalhaes, A.F. &Hambly, K. (2014). Measuring physical activity and sports participation after autologous cartilage implantation: a systematic review. *Journal of Sport Rehabilitation*. 23 (3): 171-181.

Sattelmair, J., Pertman, J., Ding, E.L., Kohl, H.W., Haskell, W. & Lee, I. (2011). Dose response between physical activity and risk of coronary heart disease a meta-analysis. *Circulation.* 124 (7): 789-795.

Savigny, P., Watson, P., Underwood, M. & Guideline Development Group (2009). Early management of persistent non-specific low back pain: summary of NICE guidance. *BMJ (Clinical research ed.).* 338 : b1805.

Scarborough, P., Bhatnagar, P., Wickramasinghe, K.K., Allender, S., Foster, C. & Rayner, M. (2011). The economic burden of ill health due to diet, physical inactivity, smoking, alcohol and obesity in the UK: an update to 2006–07 NHS costs. *Journal of public health.* 33 (4): 527-535.

Schellingerhout, J.M., Verhagen, A.P., Heymans, M.W., Koes, B.W., De Vet, H.C. & Terwee, C.B. (2012). Measurement properties of disease-specific questionnaires in patients with neck pain: a systematic review. *Quality of Life Research.* 21 (4): 659-670.

Schmitt, N.M., Schmitt, J. & Dören, M. (2009). The role of physical activity in the prevention of osteoporosis in postmenopausal women—an update. *Maturitas.* 63 (1): 34-38.

Schmitz, K.H., Courneya, K.S., Matthews, C., Demark-Wahnefried, W., Galvao, D.A., Pinto, B.M., Irwin, M.L., Wolin, K.Y., Segal, R.J., Lucia, A., Schneider, C.M., von Gruenigen, V.E., Schwartz, A.L. & American College of Sports Medicine (2010). American College of Sports Medicine roundtable on exercise guidelines for cancer survivors. *Medicine and science in sports and exercise.* 42 (7): 1409-1426.

Schmitz, K.H., Holtzman, J., Courneya, K.S., Masse, L.C., Duval, S. & Kane, R. (2005). Controlled physical activity trials in cancer survivors: a systematic review and meta-analysis. *Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology.* 14 (7): 1588-1595.

Scholtes, V.A., Terwee, C.B. & Poolman, R.W. (2011). What makes a measurement instrument valid and reliable? *Injury.* 42 (3): 236-240.

Schuit, A.J., Schouten, E.G., Westerterp, K.R. & Saris, W.H. (1997). Validity of the Physical Activity Scale for the Elderly (PASE): according to energy expenditure assessed by the doubly labeled water method. *Journal of clinical epidemiology*. 50 (5): 541-6.

Seale, J. & Rumpler, W. (1997). Comparison of energy expenditure measurements by diet records, energy intake balance, doubly labeled water and room calorimetry. *European journal of clinical nutrition.* 51 (12): 856-863.

Sebastiao, E., Gobbi, S., Chodzko-Zajko, W., Schwingel, A., Papini, C.B., Nakamura, P.M., Netto, A.V. & Kokubun, E. (2012). The International Physical

Activity Questionnaire-long form overestimates self-reported physical activity of Brazilian adults. *Public health.* 126 (11): 967-975.

Segal, N.A., Nevitt, M.C., Gross, K.D., Hietpas, J., Glass, N.A., Lewis, C.E. & Torner, J.C. (2013). The Multicenter Osteoarthritis Study: Opportunities for Rehabilitation Research. *Physical Medicine & Rehabilitation.* 5 (8): 647-654.

Segal, R.J., Reid, R.D., Courneya, K.S., Malone, S.C., Parliament, M.B., Scott, C.G., Venner, P.M., Quinney, H.A., Jones, L.W., D'Angelo, M.E. & Wells, G.A. (2003). Resistance exercise in men receiving androgen deprivation therapy for prostate cancer. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology.* 21 (9): 1653-1659.

Sellam, J. & Berenbaum, F. (2010). The role of synovitis in pathophysiology and clinical symptoms of osteoarthritis. *Nature Reviews Rheumatology.* 6 (11): 625-635.

Sigal, R.J., Kenny, G.P., Wasserman, D.H., Castaneda-Sceppa, C. & White, R.D. (2006). Physical activity/exercise and Type 2 diabetes A consensus statement from the American Diabetes Association. *Diabetes care.* 29 (6): 1433-1438.

Sim, J. & Wright, C. (2000). *Research in health care: concepts, designs and methods.* Nelson Thornes.

Silman, A.J. & Macfarlane, G.J. (2002). *Epidemiological studies: a practical guide.* Cambridge University Press, United Kingdom.

Slemenda, C., Heilman, D.K., Brandt, K.D., Katz, B.P., Mazzuca, S.A., Braunstein, E.M. & Byrd, D. (1998). Reduced quadriceps strength relative to body weight: a risk factor for knee osteoarthritis in women? *Arthritis & Rheumatism.* 41 (11): 1951-1959.

Slentz, C.A., Duscha, B.D., Johnson, J.L., Ketchum, K., Aiken, L.B., Samsa, G.P., Houmard, J.A., Bales, C.W. & Kraus, W.E. (2004). Effects of the amount of exercise on body weight, body composition, and measures of central obesity: STRRIDE—a randomized controlled study. *Archives of Internal Medicine.* 164 (1): 31-39.

Smith, S.C., Cano, S.J., Browne, J.P., Lamping, D.L., Black, N.A., van der Meulen, J., Staniszewska, S., Lewsey, J., Cairns., J. (2005) Routine use of Patient-Reported Outcome Measures (PROMs) in treatment centres: recommendations based on a review of the scientific evidence. Department of Health; London, United Kingdom

Spurr, G., Prentice, A., Murgatroyd, P., Goldberg, G., Reina, J. & Christman, N. (1988). Energy expenditure from minute-by-minute heart-rate recording: comparison with indirect calorimetry. *The American Journal of Clinical Nutrition.* 48 (3): 552-559.

Stamatakis, E., Ekelund, U. & Wareham, N.J. (2007). Temporal trends in physical activity in England: the Health Survey for England 1991 to 2004. *Preventive medicine.* 45 (6): 416-423.

Staniszewska, S., Haywood, K.L., Brett, J. & Tutton, L. (2012). Patient and public involvement in patient-reported outcome measures. *The Patient-Patient-Centered Outcomes Research.* 5 (2): 79-87.

Stehling, C., Liebl, H., Krug, R., Lane, N.E., Nevitt, M.C., Lynch, J., McCulloch, C.E. & Link, T.M. (2010). Patellar Cartilage: T2 Values and Morphologic Abnormalities at 3.0-T MR Imageing in Relation to Physical Activity in

Asymptomatic Subjects from the Osteoarthritis Initiative. *Radiology*. 254 (2): 509-520.

Steultjens, M.P., Dekker, J., van Baar, M.E., Oostendorp, R.A. & Bijlsma, J.W. (2000). Range of joint motion and disability in patients with osteoarthritis of the knee or hip. *Rheumatology.* 39 (9): 955-961.

Stevinson, C., Capstick, V., Schepansky, A., Tonkin, K., Vallance, J.K., Ladha, A.B., Steed, H., Faught, W. & Courneya, K.S. (2009). Physical activity preferences of ovarian cancer survivors. *Psycho-Oncology.* 18 (4): 422-428.

Stewart, A.L., King, A.C. & Haskell, W.L. (1993). Endurance exercise and health-related quality of life in 50-65 year-old adults. *The Gerontologist.* 33 (6): 782-789.

Strath, S.J., Kaminsky, L.A., Ainsworth, B.E., Ekelund, U., Freedson, P.S., Gary, R.A., Richardson, C.R., Smith, D.T., Swartz, A.M. & American Heart Association Physical Activity Committee of the Council on Lifestyle and Cardiometabolic Health and Cardiovascular, Exercise, Cardiac Rehabilitation and Prevention Committee of the Council on Clinical Cardiology, and Council (2013). Guide to the assessment of physical activity: Clinical and research applications: a scientific statement from the American Heart Association. *Circulation.* 128 (20): 2259-2279.

Strath, S.J., Swartz, A.M., Bassett, D.R.,Jr, O'Brien, W.L., King, G.A. & Ainsworth, B.E. (2000). Evaluation of heart rate as a method for assessing moderate intensity physical activity. *Medicine and science in sports and exercise.* 32 (9 Suppl): S465-70.

Streiner, D.L. & Norman, G.R. (2008). *Health measurement scales: a practical guide to their development and use.* Oxford university press; Oxford, United Kingdom.

Strine, T., Mokdad, A., Balluz, L., Gonzalez, O., Crider, R., Berry, J. & Kroenke, K. (2008). Depression and anxiety in the United States: findings from the 2006 behavioral risk factor surveillance system. *Psychiatric Services.* 59 (12): 1383-1390.

Stovitz, S.D., VanWormer, J.J., Center, B.A. & Bremer, K.L. (2005). Pedometers as a means to increase ambulatory activity for patients seen at a family medicine clinic. *The Journal of the American Board of Family Practice / American Board of Family Practice.* 18 (5): 335-343.

Stubbs, B., Hurley, M. & Smith, T. (2014). What are the factors that influence physical activity participation in adults with knee and hip osteoarthritis? A systematic review of physical activity correlates. *Clinical rehabilitation.* 29 (1): 80-94.

Sun, K., Song, J., Lee, J., Chang, R.W., Eaton, C.B., Ehrlich-Jones, L., Kwoh, K.C., Manheim, L.M., Semanik, P.A. & Sharma, L. (2013). Relationship of meeting physical activity guidelines with health-related utility. *Arthritis care & research.*

Svege, I., Kolle, E. & Risberg, M.A. (2012). Reliability and validity of the PA Scale for the Elderly (PASE) in patients with hip osteoarthritis. *BMC Musculoskeletal Disorders*. 13(1), 26.

Taraldsen, K., Chastin, S.F., Riphagen, I.I., Vereijken, B. & Helbostad, J.L. (2012). Physical activity monitoring by use of accelerometer-based body-worn sensors in

older adults: a systematic literature review of current knowledge and applications. *Maturitas.* 71 (1): 13-19.

Taylor, A.M., Phillips, K., Patel, K.V., Turk, D.C., Dworkin, R.H., Beaton, D., Clauw, D.J., Gignac, M.A., Markman, J.D. & Williams, D.A. (2016). Assessment of physical function and participation in chronic pain clinical trials: IMMPACT/OMERACT recommendations. *Pain*. 157 (9): 1836-1850.

Taylor, R.S., Brown, A., Ebrahim, S., Jolliffe, J., Noorani, H., Rees, K., Skidmore, B., Stone, J.A., Thompson, D.R. & Oldridge, N. (2004). Exercise-based rehabilitation for patients with coronary heart disease: systematic review and meta-analysis of randomized controlled trials. *The American Journal of Medicine.* 116 (10): 682-692.

Terwee, C.B., Bot, S.D., de Boer, M.R., van der Windt, D., Knol, D.L., Dekker, J., Bouter, L.M. & de Vet, H.C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of clinical epidemiology.* 60 (1): 34-42.

Terwee, C.B., Bouwmeester, W., van Elsland, S.L., de Vet, H.C. & Dekker, J. (2011). Instruments to assess physical activity in patients with osteoarthritis of the hip or knee: a systematic review of measurement properties. *Osteoarthritis and cartilage.* 19 (6): 620-633.

Terwee, C., Dekker, F., Wiersinga, W., Prummel, M. & Bossuyt, P. (2003). On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. *Quality of Life Research.* 12 (4): 349-362.

Terwee, C.B., Jansma, E.P., Riphagen, I.I. & de Vet, H.C. (2009). Development of a methodological PubMed search filter for finding studies on measurement

properties of measurement instruments. *Quality of Life Research.* 18 (8): 1115-1123.

Terwee, C.B., Mokkink, L.B., van Poppel, M.N., Chinapaw, M.J., van Mechelen, W. & de Vet, H.C. (2010). Qualitative attributes and measurement properties of physical activity questionnaires: a checklist. *Sports medicine.* 40 (7): 525-537.

Tessier, S., Vuillemin, A., Bertrais, S., Boini, S., Le Bihan, E., Oppert, J., Hercberg, S., Guillemin, F. & Briançon, S. (2007). Association between leisure-time physical activity and health-related quality of life changes over time. *Preventive medicine.* 44 (3): 202-208.

Thomas, K.S., Muir, K.R., DHerty, M., Jones, A.C., O'Reilly, S.C. & Bassey, E.J. (2002). Home based exercise programme for knee pain and knee osteoarthritis: randomised controlled trial. *BMJ (Clinical research ed.).* 325 (7367): 752.

Thomas, S.G., Pagura, S.M. & Kennedy, D. (2003). Physical activity and its relationship to physical performance in patients with end stage knee osteoarthritis. *Journal of Orthopaedic & Sports Physical Therapy.* 33 (12): 745-754.

Thomas, E., Peat, G., Harris, L., Wilkie, R. & Croft, P.R. (2004). The prevalence of pain and pain interference in a general population of older adults: cross-sectional findings from the North Staffordshire Osteoarthritis Project (NorStOP). *Pain.* 110 (1): 361-368.

Tipton, C.M. (2008). Susruta of India, an unrecognized contributor to the history of exercise physiology. *Journal of applied physiology.* 104 (6): 1553-1556.

Tomey, K.M. & Sowers, M.R. (2009). Assessment of physical functioning: a conceptual model encompassing environmental factors and individual compensation strategies. *Physical Therapy*. 89 (7): 705-714.

Tomioka, K., Iwamoto, J., Saeki, K. & Okamoto, N. (2011). Reliability and validity of the International Physical Activity Questionnaire (IPAQ) in elderly adults: the Fujiwara-kyo Study. *Journal of Epidemiology*. 21 (6): 459-65.

Tore Gran, J. (2003). The epidemiology of chronic generalized musculoskeletal pain. *Best Practice & Research Clinical Rheumatology.* 17 (4): 547-561.

Tran, D.V., Lee, A.H., Au, T.B., Nguyen, C.T. & Hoang, D.V. (2013). Reliability and validity of the International Physical Activity Questionnaire-Short Form for older adults in Vietnam. Health Promotion Journal of Australia. 24 (2): 126-131.

Tremblay, M.S., Warburton, D.E., Janssen, I., Paterson, D.H., Latimer, A.E., Rhodes, R.E., Kho, M.E., Hicks, A., LeBlanc, A.G. & Zehr, L. (2011). New Canadian physical activity guidelines. *Applied Physiology, Nutrition, and Metabolism.* 36 (1): 36-46.

Troiano, R.P., Berrigan, D., Dodd, K.W., Masse, L.C., Tilert, T. & McDowell, M. (2008). Physical activity in the United States measured by accelerometer. *Medicine and science in sports and exercise.* 40 (1): 181.

Trost, S.G. & O'Neil, M. (2014). Clinical use of objective measures of physical activity. *British journal of sports medicine.* 48 (3): 178-181.

Tudor-Locke, C.E. & Myers, A.M. (2001). Challenges and opportunities for measuring physical activity in sedentary adults. *Sports Medicine.* 31 (2): 91-100.

Tugwell, P., Boers, M., Brooks, P., Simon, L., Strand, V. & Idzerda, L. (2007). OMERACT: an international initiative to improve outcome measurement in rheumatology. *Trials.* 8 (1): 38.

Tully, M.A., Panter, J. & Ogilvie, D. (2014). Individual Characteristics Associated with Mismatches between Self-Reported and Accelerometer-Measured Physical Activity. *PloS one.* 9 (6): e99636.

US Department of Health and Human Services (2008). *2008 Physical Activity Guidelines for Americans.* Atlanta, Georgia, USA: US Department of Health and Human Services.

Uthman, O.A., van der Windt, Danielle A, Jordan, J.L., Dziedzic, K.S., Healey, E.L., Peat, G.M. & Foster, N.E. (2013). Exercise for lower limb osteoarthritis: systematic review incorporating trial sequential analysis and network meta-analysis. *BMJ: British Medical Journal.* 347.

University of York Centre for Reviews, and Dissemination (2009). *Systematic reviews: CRD's guidance for undertaking reviews in health care.* York: CRD, University of York, United Kingdom.

Van Agt, H.M., Essink-Bot, M., Krabbe, P.F. & Bonsel, G.J. (1994). Test-retest reliability of health state valuations collected with the EuroQol questionnaire. *Social science & medicine.* 39 (11): 1537-1544.

Van Baar, M.E., Dekker, J., Oostendorp, R.A., Bijl, D., Voorn, T.B. & Bijlsma, J.W. (2001). Effectiveness of exercise in patients with osteoarthritis of hip or knee: nine months' follow up. *Annals of the Rheumatic Diseases*. 60 (12): 1123-1130.

Van der Bij, Akke K, Laurant, M.G. & Wensing, M. (2002). Effectiveness of physical activity interventions for older adults. *American Journal of Preventive Medicine.* 22 (2): 120-133.

Van der Kraan, P. (2012). Osteoarthritis year 2012 in review: biology. *Osteoarthritis and Cartilage.*

Van Holle, V., De Bourdeaudhuij, I., Deforche, B., Van Cauwenberg, J. & Van Dyck, D. (2015). Assessment of physical activity in older Belgian adults: validity and reliability of an adapted interview version of the long International Physical Activity Questionnaire (IPAQ-L). *BMC Public Health.* 15: 433.

Van Poppel, M.N., Chinapaw, M.J., Mokkink, L.B., van Mechelen, W. & Terwee, C.B. (2010). Physical activity questionnaires for adults: a systematic review of measurement properties. *Sports medicine.* 40 (7): 565-600.

Valdes, A.M., DHerty, M. & Spector, T.D. (2008). The additive effect of individual genes in predicting risk of knee osteoarthritis. *Annals of the Rheumatic Diseases.* 67 (1): 124-127.

Vaughan, K. & Miller, W.C. (2013). Validity and reliability of the Chinese translation of the Physical Activity Scale for the Elderly (PASE). *Disability & Rehabilitation.* 35 (3): 191-197.

Verweij, L.M., Schoor, N.M.v., Deeg, D.J., Dekker, J. & Visser, M. (2009). Physical activity and incident clinical knee osteoarthritis in older adults. *Arthritis & Rheumatism*. 61 (2): 152-157.

Vickers, A.J. (2005). Parametric versus non-parametric statistics in the analysis of randomized trials with non-normally distributed data. BMC medical research methodology. 5 : 35.

Vina, J., Sanchis-Gomar, F., Martinez-Bello, V. & Gomez-Cabrera, M. (2012). Exercise acts as a drug; the pharmacological benefits of exercise. *British journal of pharmacology.* 167 (1): 1-12.

Voorrips, L.E., Ravelli, A.C., Dongelmans, P.C., Deurenberg, P. & Van, W.A.,WA (1991). A physical activity questionnaire for the elderly. *Medicine & Science in Sports & Exercise*. 23 (8): 974-9.

Vuillemin, A., Boini, S., Bertrais, S., Tessier, S., Oppert, J., Hercberg, S., Guillemin, F. & Briançon, S. (2005). Leisure time physical activity and health-related quality of life. *Preventive medicine.* 41 (2): 562-569.

Waddell, G. & Burton, A.K. (2001). Occupational health guidelines for the management of low back pain at work: evidence review. *Occupational medicine (Oxford, England).* 51 (2): 124-135.

Wagenmakers, R., Stevens, M., Zijlstra, W., Jacobs, M.L., Akker-Scheek, I.v.d., Groothoff, J.W. & Bulstra, S.K. (2008). Habitual physical activity behavior of patients after primary total hip arthroplasty. *Physical Therapy*. 88 (9): 1039-1048.

Wang, Y.C., McPherson, K., Marsh, T., Gortmaker, S.L. & Brown, M. (2011). Health and economic burden of the projected obesity trends in the USA and the UK. *The Lancet.* 378 (9793): 815-825.

Wang, Q., Rozelle, A.L., Lepus, C.M., Scanzello, C.R., Song, J.J., Larsen, D.M., Crish, J.F., Bebek, G., Ritter, S.Y. & Lindstrom, T.M. (2011). Identification of a central role for complement in osteoarthritis. *Nature medicine.* 17 (12): 1674-1679.

Warburton, D.E., Nicol, C.W. & Bredin, S.S. (2006). Health benefits of physical activity: the evidence. *CMAJ : Canadian Medical Association journal.* 174 (6): 801-809.

Ware Jr, J.E. & Gandek, B. (1998). Overview of the SF-36 health survey and the international quality of life assessment (IQOLA) project. *Journal of clinical epidemiology.* 51 (11): 903-912.

Ware Jr, J.E., Kosinski, M. & Keller, S.D. (1996). A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Medical care.* 34 (3): 220-233.

Warren, J.M., Ekelund, U., Besson, H., Mezzani, A., Geladas, N. & Vanhees, L. (2010). Assessment of physical activity–a review of methodologies with reference to epidemiological research: a report of the exercise physiology section of the European Association of Cardiovascular Prevention and Rehabilitation. *European Journal of Cardiovascular Prevention & Rehabilitation.* 17 (2): 127-139.

Washburn, R.A. & Ficker, J.L. (1999). Physical Activity Scale for the Elderly (PASE): the relationship with activity measured by a portable accelerometer. *Journal of Sports Medicine & Physical Fitness*. 39 (4): 336-40.

Washburn, R.A., Smith, K.W., Jette, A.M. & Janney, C.A. (1993). The Physical Activity Scale for the Elderly (PASE): development and evaluation. *Journal of clinical epidemiology.* 46 (2): 153-162.

Watz, H., Waschki, B., Meyer, T. & Magnussen, H. (2009). Physical activity in patients with COPD. *European Respiratory Journal.* 33 (2): 262-272.

Weir, J.P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of strength and conditioning research.* 19 (1): 231-240.

Weller, I.M. & Kunz, M. (2007). Physical activity and pain following total hip arthroplasty. *Physiotherapy*. 93 (1): 23-29.

Weldam, S.W., Schuurmans, M.J., Liu, R. & Lammers, J.J. (2013). Evaluation of Quality of Life instruments for use in COPD care and research: a systematic review. International journal of nursing studies. 50 (5): 688-707.

Wen, C.P., Wai, J.P.M., Tsai, M.K., Yang, Y.C., Cheng, T.Y.D., Lee, M., Chan, H.T., Tsao, C.K., Tsai, S.P. & Wu, X. (2011). Minimum amount of physical activity for reduced mortality and extended life expectancy: a prospective cohort study. *The Lancet.* 378 (9798): 1244-1253.

Westerterp, K.R. (2009). Assessment of physical activity: a critical appraisal. *European journal of applied physiology.* 105 (6): 823-828.

Wendel-Vos, G., Schuit, A., Tijhuis, M. & Kromhout, D. (2004). Leisure time physical activity and health-related quality of life: cross-sectional and longitudinal associations. *Quality of Life Research.* 13 (3): 667-677.

Whelton, S.P., Chin, A., Xin, X. & He, J. (2002). Effect of aerobic exercise on blood pressure a meta-analysis of randomized, controlled trials. *Annals of Internal Medicine.* 136 (7): 493-503.

White, D.K., Tudor-Locke, C., Felson, D.T., Gross, K.D., Niu, J., Nevitt, M., Lewis, C.E., Torner, J. & Neogi, T. (2013). Do radiographic disease and pain account for why people with or at high risk of knee osteoarthritis do not meet physical activity guidelines? *Arthritis & Rheumatism.* 65 (1): 139-147.

Whitley, E. & Ball, J. (2002). Statistics review 6: Nonparametric methods. *Critical Care.* 6 (6): 1-5.

Wijndaele, K., Duvigneaud, N., Matton, L., Duquet, W., Delecluse, C., Thomis, M., Beunen, G., Lefevre, J. & Philippaerts, R. (2007). Sedentary behaviour, physical

activity and a continuous metabolic syndrome risk score in adults. *European journal of clinical nutrition.* 63 (3): 421-429.

Williamson, P.R., Altman, D.G., Blazeby, J.M., Clarke, M., Devane, D., Gargon, E. & Tugwell, P. (2012). Developing core outcome sets for clinical trials: issues to consider. Trials. 13 (132): 1-8.

Winkler, E., Waters, L., Eakin, E., Fjeldsoe, B., Owen, N. & Reeves, M. (2013). Is measurement error altered by participation in a physical activity intervention?. *Medicine & Science in Sports & Exercise.* 45 (5): 1004-1011.

Wollmerstedt, N., Noeth, U., Ince, A., Ackermann, H., Martell, J.M. & Hendrich, C. (2010). The Daily Activity Questionnaire A Novel Questionnaire to Assess Patient Activity After Total Hip Arthroplasty. *Journal of Arthroplasty*. 25 (3): 475-480.

Woolf, A.D. & Pfleger, B. (2003). Burden of major musculoskeletal conditions. *Bulletin of the World Health Organization.* 81 (9): 646-656.

World Health Organisation (WHO) (2010). Global recommendations on physical activity for health. Geneva, Switzerland: World Health Organisation.

Wright, J.G. & Feinsten, A.R. (1992). A comparative contrast of clinimetric and psychometric methods for constructing indexes and rating scales. *Journal of clinical epidemiology.* 45 (11): 1201-1218.

Yorston, L.C., Kolt, G.S. & Rosenkranz, R.R. (2012). Physical activity and physical function in older adults: the 45 and up study. *Journal of the American Geriatrics Society.* 60 (4): 719-725.

Zelman, D.C., Dukes, E., Brandenburg, N., Bostrom, A. & Gore, M. (2005). Identification of cut-points for mild, moderate and severe pain due to diabetic peripheral neuropathy. *Pain.* 115 (1): 29-36.

Zeni, J.A., Jr & Higginson, J.S. (2009). Dynamic knee joint stiffness in subjects with a progressive increase in severity of knee osteoarthritis. *Clinical biomechanics.* 24 (4): 366-371.

Zhang, Y. & Jordan, J.M. (2010). Epidemiology of osteoarthritis. *Clinics in geriatric medicine.* 26 (3): 355-369.

Zigmond, A.S. &Snaith, R.P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica.* 67 (6): 361-370.

Zochling, J., van der Heijde, D., Burgos-Vargas, R., Collantes, E., Davis, J.C.,Jr, Dijkmans, B., Dougados, M., Geher, P., Inman, R.D., Khan, M.A., Kvien, T.K., Leirisalo-Repo, M., Olivieri, I., Pavelka, K., Sieper, J., Stucki, G., Sturrock, R.D., van der Linden, S., Wendling, D., Bohm, H., van Royen, B.J., Braun, J., 'ASsessment in AS' international working group & European League Against Rheumatism (2006). ASAS/EULAR recommendations for the management of ankylosing spondylitis. *Annals of the Rheumatic Diseases.* 65 (4): 442-452.

12 Appendix

Appendix 1.1 MOSAICS protocol paper (With permission)

Dziedzic et al. Implementation Science 2014, 9:95
http://www.implementationscience.com/content/9/1/95

IMPLEMENTATION SCIENCE

STUDY PROTOCOL Open Access

Implementing the NICE osteoarthritis guidelines: a mixed methods study and cluster randomised trial of a model osteoarthritis consultation in primary care - the Management of OsteoArthritis In Consultations (MOSAICS) study protocol

Krysia S Dziedzic[1], Emma L Healey[1*], Mark Porcheret[1], Bie Nio Ong[1], Chris J Main[1], Kelvin P Jordan[1], Martyn Lewis[1], John J Edwards[1], Clare Jinks[1], Andrew Morden[1], Gretl A McHugh[2], Sarah Ryan[3], Andrew Finney[1], Sue Jowett[1,4], Raymond Oppong[1,4], Ebenezer Afolabi[1], Angela Pushpa-Rajah[5], June Handy[1], Kris Clarkson[1], Elizabeth Mason[1], Tracy Whitehurst[1], Rhian W Hughes[1], Peter R Croft[1] and Elaine M Hay[1]

Abstract

Background: There is as yet no evidence on the feasibility of implementing recommendations from the National Institute of Health and Care Excellence (NICE) osteoarthritis (OA) guidelines in primary care, or of the effect these recommendations have on the condition. The primary aim of this study is to determine the clinical and cost effectiveness of a model OA consultation (MOAC), implementing the core recommendations from the NICE OA guidelines in primary care. Secondary aims are to investigate the impact, feasibility and acceptability of the MOAC intervention; to develop and evaluate a training package for management of OA by general practitioners (GPs) and practice nurses; test the feasibility of deriving 'quality markers' of OA management using a new consultation template and medical record review; and describe the uptake of core NICE OA recommendations in participants aged 45 years and over with joint pain.

Design: A mixed methods study with a nested cluster randomised controlled trial.

Method: This study was developed according to a defined theoretical framework (the Whole System Informing Self-management Engagement). An overarching model (the Normalisation Process Theory) will be employed to undertake a comprehensive 'whole-system' evaluation of the processes and outcomes of implementing the MOAC intervention. The primary outcome is general physical health (Short Form-12 Physical component score [PCS]) (Ware 1996). The impact, acceptability and feasibility of the MOAC intervention at practice level will be assessed by comparing intervention and control practices using a Quality Indicators template and medical record review. Impact and acceptability of the intervention for patients will be assessed via self-completed outcome measures and semi-structured interviews. The impact, acceptability and feasibility of the MOAC intervention and training for GPs and practice nurses will be evaluated using a variety of methods including questionnaires, semi-structured interviews, and observations.

(Continued on next page)

* Correspondence: e.healey@keele.ac.uk
[1]Arthritis Research UK Primary Care Centre, Keele University, Keele, Staffordshire ST5 5BG, UK
Full list of author information is available at the end of the article

(Continued from previous page)

Discussion: The main output from the study will be to determine whether the MOAC intervention is clinically and cost effective. Additional outputs will be the development of the MOAC for patients consulting with joint pain in primary care, training and educational materials, and resources for patients and professionals regarding supported self-management and uptake of NICE guidance.

Trial registration: ISRCTN number: ISRCTN06984617.

Keywords: Osteoarthritis, General practice, Implementation, Primary care, NICE guidelines, Self-management

Background

There is a perception that osteoarthritis (OA) is a 'natural' part of ageing and there are limited interventions available [1,2]. Whilst there are many published guidelines on the treatment of OA [3-10] there is a gap between the care that is recommended and the care that patients receive [11,12]. NICE have recommended that all patients with OA should be offered three core treatments when they first present in primary care (see Figure 1): education and access to information; advice on local muscle strengthening exercise and general aerobic fitness; and, if appropriate, advice on losing weight [13].

For patients and healthcare professionals (HCPs) in primary care, the clear message that emerges from the guidelines is that there is a range of simple interventions for which there is evidence of clinical effectiveness. By contrast, there is clear evidence that the core aspects of assessment and management of OA as currently delivered in primary care do not meet the recommendations of these guidelines [11]. Two recent Delphi consensus exercises have been conducted to address the limited research evidence on the content of a model OA consultation during which the core package of care could be delivered [14,15]. However, there is as yet no evidence on the feasibility of using this model consultation as a way of implementing NICE core OA treatments in primary care and the effect of this support for self-management on the course and impact of the condition.

The importance of self-management for long term illnesses and of professional support for self-management are emphasised in NHS policy [16]. Studies have shown that among patients consulting for knee OA, core treatments were mostly self- rather than doctor-initiated [1,11]. Previous research suggests that patients need more information about OA to enable them to manage their condition more effectively [17]. Lay ideas of self-management include minimising the impact of conditions [18,19], maintaining a sense of 'normality' [18,20], and maintaining social roles and obligations [21,22].

The way in which complex interventions, such as supported self-management approaches, are developed and then become embedded in routine clinical practice needs to be supported by theoretical models and evaluated appropriately [23]. A specific model of support for self-management called the 'Whole system Informing Self-management Engagement' (WISE) model [24] is predicated upon the argument that for self-management support

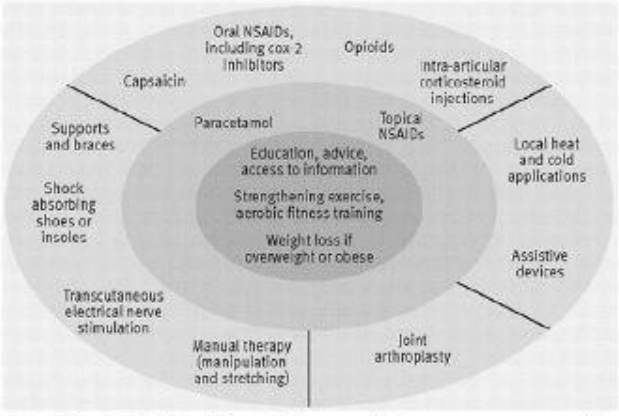


Figure 1 National Institute for Health and Clinical Excellence (NICE) osteoarthritis treatment recommendations (Conaghan et al. [10]). Reproduced from, Care and management of osteoarthritis in adults: summary of NICE guidance, Conaghan P, Dickson J, Grant R, 336, 502-503, 2008, with permission from BMJ Publishing Group Ltd.

to be effective, it requires the understanding and incorporation of the patient agenda into the consultation, including what they are already doing. A whole systems approach is needed, which engages with practitioners and service organisations as well as the patient [24]. The WISE model has been used in a number of long-term illnesses [25-27], and envisages informed patients receiving support and guidance from trained practitioners who are working within a healthcare system that is geared up to be responsive to patients' needs.

Michie and colleagues have produced a synthesis of psychological theories to enable design and implementation of behaviour change interventions [28-30], and Grol and colleagues have developed a useful framework for translating evidence into practice [31]. The Calgary-Cambridge framework has specifically been developed to enhance consultation skills [32]. The Normalisation Process Theory (NPT) [33,34] is a sociological theory concerned with understanding and evaluating how complex interventions become embedded in routine clinical practice. These approaches will be integrated to develop the model OA consultation (MOAC) intervention and the HCP training in this study.

This study is focused on determining the current management of OA in consultations and whether supported self-management, delivered in a model OA consultation, could offer a clinically practicable approach to implementing the core NICE recommendations. We will undertake a cluster randomised controlled trial (RCT) to evaluate the whole systems approach and to determine the clinical and cost effectiveness of the MOAC intervention. The protocol has been reported using the SPIRIT recommendations [35,36].

Aims

The primary aim of the study is to:

1. Determine the clinical and cost effectiveness of the MOAC intervention in patients with OA.

Secondary aims are to:

1. Describe the uptake of core NICE OA recommendations in participants aged 45 years and over with joint pain;
2. Test the feasibility of deriving 'quality markers' of OA management using a new consultation template and medical record review;
3. Develop and evaluate a training package for management of OA by general practitioners (GPs) and practice nurses;
4. Investigate the impact, feasibility and acceptability of the MOAC intervention.

Design

The MOSAICS study is a mixed methods study with a nested cluster RCT based in eight general practices, comprising of four components (see Figure 2).

Participants

HCPs and their respective practice populations from eight general practices will be invited to participate and randomly allocated to two clusters: 'intervention' and 'control' practices. The practice populations of the practices recruited, aged 45 years and over, will form the sampling frame for the overall study. The eligibility criteria for the study are described in Table 1. Resources to support primary care engagement are described in Additional file 1.

Ethics approval

The study has been approved by the North West 1 Research Ethics Committee, Cheshire (REC reference: 10/H1017/76).

Consent

Informed consent will be obtained from all HCPs and patients participating in the study (see Additional file 2).

Components of the MOSAICS study

1. The cluster randomised controlled trial
2. Population survey
3. Consultation template and medical record review
4. Evaluation of the MOAC intervention and the training

Study component 1: The cluster RCT

This will be a two-arm prospective pragmatic cluster RCT based in eight general practices in the North West midlands and Cheshire, UK. This trial will test the hypothesis that the MOAC intervention is superior to control and will be a parallel group longitudinal design with repeated measures. Practices will be randomly allocated to two clusters: 'intervention' and 'control' practices. Stratified block randomisation will be performed with the practices stratified into four largest and four smallest practices by list size. In each stratum, the practices will be randomly allocated by a computerised random number generator using blocking with two practices per study group.

Recruitment to the trial will occur over a six-month period. All practices will use the OA template and will be given standard information on OA for patients and healthcare professionals published by Arthritis Research UK. In four practices, patients will receive the MOAC intervention, while the other four practices will continue to provide usual care. GPs in the intervention arm will be offered training for the initial consultation (MOAC 1), and practice nurses

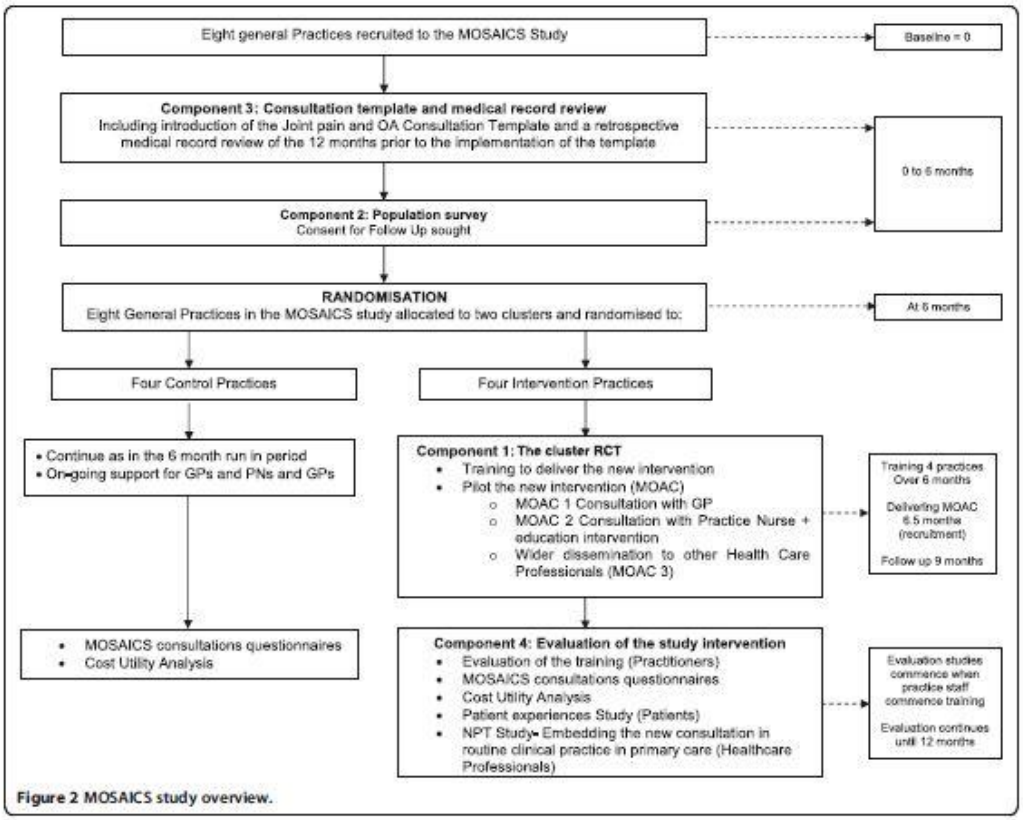


Figure 2 MOSAICS study overview.

will be offered comprehensive training for follow-up consultations (MOAC 2), to deliver supported self-management of OA. The control practices will have the opportunity to receive the training at the end of data collection.

The intervention

The MOAC intervention

In order to identify the content of the MOAC intervention, two consensus exercises with GPs, lay participants, practice nurses, community pharmacists, and allied health professionals (occupational therapists, physiotherapists, podiatrists) have been conducted [14,15]. Using the findings of these exercises and theoretical models to guide self-management (WISE) and support patient behaviour change, a three-stage consultation has been proposed. The MOAC intervention comprises of MOAC 1 (initial consultation with GP), MOAC 2 (follow-up visits with practice nurse) [37], and MOAC 3 (dissemination to other HCPs). Consultations will be recorded in the general practices through the use of the joint pain and OA consultation template (Additional file 3 provides the full details of the intervention).

The OA Guidebook

The Arthritis Research UK Primary Care Centre at Keele University has developed a guidebook for patients and professionals to use as an aid to support self-management for OA, and this will be used in MOAC [38]. The OA Guidebook contains both evidence-based biomedical information and lay experiential knowledge about the nature of OA, how it is diagnosed and treated, and the ways in which people who have OA continue to keep going in their everyday lives. Together these will give patients insight into the rationale for the advice and treatments offered in MOAC 1, 2 and 3 and what, practically, people who have OA have found helpful in learning to live with the condition. The patient guidebook will be a useful way to reinforce verbal information given by HCPs, to act as a resource to consult if queries or uncertainties arise in the future, to help

Table 1 MOSAICS study eligibility criteria

Eligibility criteria	
General practices and health-care professionals	• Member of the Central England PCRN or a Keele Research Network Practice
	• At least two GPs willing to undertake the study as per protocol, *i.e.*, act as a control or intervention practice
	• Willing, and able, to allow one (or for preference two – to allow for cross cover) of their practice nurses to be trained to deliver the MOAC 2 clinics
	• Able to physically accommodate the nurse clinics in the practice
	• Uses the EMIS computerised consultation system
	• Nurses and GPs consenting to follow up by the MOSAICS study team
	• GPs willing to be trained to carry out the MOAC 1 consultations
	• Nurses willing to be trained to carry out the MOAC 2 clinics
	• Nurses who consent to being observed and audio recorded in MOAC 2 clinics
***Patients**	Inclusion Criteria:
	• Males and Females
	• 45 years and over
	• Registered with a MOSAICS study practice
	• Consenting to further contact from the study team and medical record review (consent sought as part of the Patient Population Survey)
	Exclusion Criteria:
	• Excluded via GP screen of practice list
	• Unable to give fully informed consent, *e.g.*, learning difficulties or dementia
	• Resident in a care or nursing home
	• History of serious disease, *e.g.*, malignancy, terminal illness
	• Unable to consult in the general practice
	• Flagged as excluded from research in that practice

EMIS = Egton Medical Information System; GP = General Practitioners; MOAC = Model Osteoarthritis Consultation; MOSAICS = Management of Osteoarthritis in consultations; OA = Osteoarthritis; PCRN = Primary Care Research Network.
*All patients registered with MOSAICS practices randomised to the intervention arm of the study had access to the MOAC 1 and MOAC 2 clinics, including those who did not consent to be followed up, as these ran as part of normal clinical care within each intervention practice.

structure MOAC 2 consultations, and to prompt questions to ask HCPs.

The training

Training and educational packages will be developed for GPs and practice nurses by drawing on the work of Michie et al. [29,30]. GPs will receive training on how to deliver the initial consultation (MOAC 1) for a new or established patient consulting with OA and the procedure for referring to a practice nurse for a follow-up OA consultation (MOAC 2). Practice nurses will receive training in how to support and enable patients to self-manage OA, using a patient-centred approach, an OA guidebook, goal setting, pain management and the core NICE recommendations (information and advice, strengthening exercise and aerobic fitness training, and weight management). Nurses will also receive training in how to complete the MOAC 2 case report form (CRF), which will be completed for each patient attending the MOAC 2 consultations. Full details of these training packages will be published separately.

For MOAC 3, members of the wider multidisciplinary team linked to the intervention practices (*e.g.*, physiotherapists, occupational therapists, podiatrists, pharmacists) will be invited to workshops to increase awareness of the study and its aims.

Sample size

Published research in musculoskeletal disorders has estimated a minimal clinically important difference (MCID) from 2 to 4 points for the SF-36 version 2 Physical Component Summary (PCS) subscale [39], which has a normalised population standard deviation (SD) of 10. Hence, a difference between groups of 0.3 of an effect size is considered a minimum clinically important difference (MCID) threshold for the purposes of demonstrating superiority in the MOSAICS RCT.

In total, 500 patients will need to be recruited in expectation that 400 will provide data at six months. A total of 400 participants (200 to each arm) will ensure 90% power to detect at least the effect size of 0.3 at the primary time point of six months follow-up given a 5% two-tailed significance level. Randomisation is by practice, so this sample size calculation was inflated to correct for an intracluster correlation coefficient (adjusted ICC of 0.005), varying practice size recruitment was taken into account (including coefficient of variation of 0.5 as per POST trial [ISRCTN40721988]) and inclusive of ($\times 0.67$) and $\times 1.25$ respective adjustments for repeated-measures design and 20% dropout allowance.

Patient level evaluation

Patient consultation questionnaire

Eligible participants who consented to medical record review and further contact in the initial baseline patient population survey (see component 2), and who subsequently consult with joint pain in one of the four intervention practices will be identified via fortnightly electronic searches of consultation data and receive a consultation questionnaire asking about their consultation experience with the MOAC intervention. This will also be

administered to eligible participants in the four control practices, to enable comparison of patient reported clinical outcomes.

The primary clinical outcome measure for the cluster RCT is the SF12 PCS [40]. The primary follow-up time point is six months. Key secondary outcomes include the Arthritis Self-Efficacy pain subscale [41] and the OMERACT/OARSI responder criteria [42]. Other outcome measures collected in the trial can be found in (Table 2). Participants will be followed up at 3, 6 and 12 months after consultation to determine short, medium and long term behaviours and patterns of the uptake of core OA treatments. Process outcome measures will also be collected within the study (*e.g.*, achieving Quality Indicators of care).

The trial statistician will be kept blind to the practice allocation until after the analysis of the primary and secondary outcomes (blinding will be broken for the per protocol analysis) [43].

Objectives: Study component 1

1a. To determine the clinical and cost effectiveness of the MOAC intervention in four general practices compared with four control practices using patient reported outcome measures.
1b. To train healthcare professionals (GPs and practice nurses) to deliver the MOAC intervention.
1c. To implement the MOAC intervention to deliver supported self-management for adults with OA within routine primary care.

Trial status

Eight general practices have been recruited, and the intervention has been developed and delivered in the four intervention practices. We expect follow-up data collection to be complete in 2014.

Study component 2: Population survey

A cross sectional population survey will be mailed to an estimated sample of 30,000 adults aged 45 years and over registered in the eight participating general practices participating in the study. After exclusions and based on previous similar studies from the Research Centre, we anticipate a sample of 9,600 with self-reported joint pain.

The population survey will use a 2-stage mailing process based on established procedures conducted in the Arthritis Research UK Primary Care Centre. Eligible participants will be sent a letter of invitation to take part in the survey, information about the study, and the population survey. This survey will collect demographic and work-related data and ask questions regarding general and psychological health, physical activity, joint pain in the last 12 months, consultation behaviour and the management of their joint pain (knee, hip, foot, hand) (the full list of outcome measures can be found in Table 3). Consent will be sought for further involvement in the MOSAICS study and for allowing access to their medical records. Individuals excluded by the GP or contacting the research team and not wishing to take part in the study will be tagged in the practices as exclusions and will not be contacted again for this study. After two weeks, non-responders will be sent a reminder survey and letter.

Those that respond to the population survey and agree further contact will be approached for data collection regarding the cluster RCT if they consult with joint pain during the recruitment period. The key reason for doing this is to minimise selection bias in the trial by separating the process of consent from the intervention.

Objectives: Study component 2

2a. To describe pain severity, general health, psychological status, and uptake of core treatments for OA recommended by NICE in participants aged 45 years and over with joint pain (hip, knee, hand, foot).
2b. To identify a population within the practices that agrees to further contact and medical record review and can therefore be approached for data collection for the cluster RCT.

Study component 3: Consultation template and medical record review

This component will collect anonymised practice-level data to describe the management of OA in primary care via retrospective and prospective medical record review in the eight practices recruited to the study. The OA consultation template will record specific Quality Indicators for the core management of OA which are not routinely recorded in practice. These indicators were established through a systematic review of Quality Indicators of OA [48]. These include pain and functional impairment, provision of information, advice about exercise and weight loss, and advice about pharmacological management such as use of paracetamol and topical non-steroidal anti-inflammatory drugs (NSAIDs).

A download of routine consultation, prescription and other management data related to joint pain and OA will be obtained for the preceding 12 months before the introduction of the template. Data regarding certain co-morbidities, identified as having implications for prescribing behaviours in the management of joint pain and OA, will also be obtained for three years preceding the introduction of the template.

The OA consultation template will then be introduced in all of the eight practices at baseline (six months prior to randomisation of the practices for the cluster RCT described in component 1). Training regarding use of

Table 2 Consultation survey measures

Data Collection	Measurement Scale	Time points			
		Baseline	3 months	6 Months	12 Months
Demographic Information					
Age	Years	✓	✓	✓	✓
Gender	Female/Male	✓	✓	✓	✓
Weight	Stones and lbs or Kilograms	✓	✓	✓	✓
Work related Questions					
Current/most recent job title	Free Text	✓	✓	✓	✓
Currently in a paid Job	Yes/No/Retired	✓	✓	✓	✓
Typical working week	Working full time (30 hours or more per week)/ Working part time (29 hours or less per week)/	✓	✓	✓	✓
Time off during last 6 months because of joint pain including time off to visit any health care professional	Yes/No	✓	✓	✓	✓
How many days, weeks or months were you absent from work due to joint pains in the last 6 months	Number of days/weeks/months	✓	✓	✓	✓
Consultations					
GPAQ communication sub scale [43]	8 – 56	✓			
GPAQ for Nurses communication sub scale (modified GPAQ [43])	8 – 56		✓		
Consulting Practice Nurse (single question)	Yes/No		✓		
Joint Pain					
Specific Joint Pain and Problems in knee/hip/hand/foot over 3 months	Yes/No	✓	✓	✓	✓
Pain intensity in the knee/hip/hand/foot over 3 months	0-10 numerical rating scale	✓	✓	✓	✓
WOMAC Physical function subscale [44]	0 - 32	✓	✓	✓	✓
AIMS 2 hand and finger function subscale [45]	0-20	✓	✓	✓	✓
Physical Activity					
IPAQ [46]	Categorical score: low, moderate, high Continuous score: MET-min per week	✓	✓	✓	✓
Walking Questions	Where do you regularly walk for reasons including health and well-being?/ Who do you regularly walk with?	✓	✓	✓	✓
PASE [47]	0 – 361	✓	✓	✓	✓
Global Assessment of change [48]	Completely recovered, much better, better, no change, worse, much worse		✓	✓	✓
General health and Well being					
PHQ8 [49]	0 – 24	✓	✓	✓	✓
GAD7 [50]	0 – 21	✓	✓	✓	✓
SF12 version 2 [40]	0 – 100	✓	✓	✓	✓
SF6D [51]	0.29-1				
EQ-5D [52]	-0.59-1	✓	✓	✓	✓
ICECAP-A version 2 [53]	4 - 20	✓	✓	✓	✓
Managing your joint problems					
Arthritis Self Efficacy pain sub scale [41]	0 - 10	✓	✓	✓	✓

Table 2 Consultation survey measures *(Continued)*

Patient generated OA Quality Indicators (Adapted from Østerås et al [54])	15 questions (Yes/No/Don't remember)	✓	✓	✓	✓
Medication/Treatment Use (adapted for joint problems from Jinks et al [55])	Simple count of strategies used	✓	✓	✓	✓
Patient Enablement (modified from Howie et al [56])	0 - 10		✓	✓	✓
Healthcare Utilisation	Self-help remedies, contact with NHS and private healthcare, over the counter medicines, prescribed medication			✓	✓

Key:
AIMS 2 = Arthritis Impact Measurement Scale; GAD 7= Generalized Anxiety Disorder 7; GP = General Practitioner; GPAQ = General Practice Assessment Questionnaire;
ICECAP-A = self-report measure of capability wellbeing for adults; IPAQ: International Physical Activity Questionnaire; MET = Metabolic Equivalent; PASE = Physical Activity Scale for the Elderly; PHQ8 = Patient Health Questionnaire 8; EQ5D = Quality of Life; SF12 = Short Form 12; SF6D = Short Form 6D; WOMAC = Western Ontario and McMaster Universities Arthritis Index.

the template as part of routine consultations for patients presenting with a working diagnosis of OA (knee, hip, hand or foot pain) will be provided for the GPs and practice nurses. When a patient aged 45 or over with joint pain consults in any of the eight practices and the GP or practice nurse enters a Read code for OA or one of a selection of joint pain Read codes (see Additional file 4), the OA consultation template will open to allow data entry. The GP or practice nurse undertaking the consultation may complete the template, or bypass the template, as appropriate (for various reasons, such as the joint pain not being related to OA). Data on routine recording of management and Quality Indicators will be collected in the template prospectively at 6 months after baseline and 21 months by downloading anonymised practice data.

Objectives: Study component 3

3a. To describe the current recording patterns and management of OA in primary care based on a 12-month retrospective download of medical records.
3b. To describe change in the recording patterns and management of OA in primary care following (i) activation of the template, and (ii) delivery of the MOAC intervention, based on prospective downloads of medical records.

Study component 4: Evaluation of the MOAC intervention and the training

This component will evaluate (i) the implementation of the MOAC intervention at the level of the service, HCPs and patients, and (ii) the training. Both qualitative and quantitative methods will be used.

Evaluation of the intervention

Practice level evaluation

Observations of a number of relevant meetings will be undertaken. The number of observations will be decided, in negotiation with each intervention practice. The intention is to observe all relevant clinical, management and research meetings to reveal how the intervention is operationalised within practice. In order to document events, observations will be structured and will use audio recordings and an observation framework to capture data for analysis. Secondary data collection will also be undertaken to determine adherence to the intervention. For example, notes of meetings in the practices, MOAC 2 CRFs, patient records, and patient feedback forms will be used to gain an understanding of how the intervention was taken up and delivered.

HCP level evaluation

In order to capture the complexity of how the intervention is received, delivered and managed by members of the primary healthcare team, a range of methods will be used. These methods include semi-structured face-to-face interviews with HCPs, structured observations of consultations, video-recorded simulated consultations, and secondary data collection (for example, minutes of meetings).

GP and practice nurse interviews and focus group

As part of the last training session, all GPs and practice nurses will be invited to participate in a focus group, to ascertain whether, and if so how, their views of OA treatment have changed since the training, and to explore what this means for their clinical practice. GPs and nurses from the intervention practices will also be invited to participate in individual telephone or face-to-face interviews after completion of the intervention. All interviews will be digitally-recorded and fully transcribed and conducted six to nine months following randomisation.

Practice nurse observations

Observations of MOAC 2 consultations will be undertaken to gain detailed insight into the consultation itself, the interaction between the nurse and patient, and the actual delivery of MOAC 2 in routine practice. Audio recordings and detailed field notes will capture data for

Table 3 Population survey outcome measures

Data Collection	Measurement Scale
Demographic	
Age	Years
Gender	Female/Male
Marital status	Married/separated/divorced/widowed/cohabiting/single
Living alone	Yes/No
Spouse/Partner cohabiting	Yes/No
Cost of living	Find it a strain to get by from week to week/Have to be careful with Money/Able to manage without much difficulty/Quite comfortably off
Height	Feet and Inches or centimetres
Weight	Stones and lbs or kilograms
Work related questions to determine occupational social class	
Employment status	Employed / Not working due to ill health / Retired/Unemployed or seeking work/ Housewife/Other
Spouse/partner's job title	Free Text
Work related questions including:	Current employment status; Job title (or previous job title if retired/unemployed); Spouse's job title (or previous job title if no longer working or deceased)
General health	
SF12 version 2 (physical and mental summary score) [40]	0 - 100
GAD7 [50]	0 - 21
EQ-5D 3 level version [52]	-0.59 - 1
Physical Activity	
STAR [58]	Three categories: 1) Physically inactive, 2) Meets the current recommendations of physical activity, 3) Insufficiently active
Joint Pain	
Specific joint pain and problems in knee/hip/hand/foot over past year	Yes/No
Pain intensity in the knee/hip/hand/foot over past month	0-10
Managing Joint Problems over the last 12 months	
Consultation with practice nurse (single question)	Yes/No
Consultation with GP (single question)	Yes/No
Medication/Treatment use (adapted for joint problems from Jinks et al [55])	Simple count of strategies used

GP = General Practitioner; GAD 7 = Generalized Anxiety Disorder 7; EQ5D = Quality of Life; SF12 = Short Form 12; STAR = Short Telephone Activity Rating.

analysis and allow verbal and non-verbal communications to be compared.

Video-recorded simulated patient consultations

To evaluate OA consultation behaviour, the GPs in the intervention practices will be invited to undertake video-recorded consultations with simulated patients. The simulated patient will take on the role of one with chronic joint problems, and the GP will be asked to conduct a consultation in which the problems are assessed and a management plan agreed upon. Video-recorded consultations will be undertaken (i) before the training (video 1), (ii) during training (video 2), (iii) one month after training (video 3), and (iv) six months after the training (video 4). Two video-recorded consultations (videos 1 and 2) will be used as part of the training to enable GPs to reflect on their consultation behaviours and video-recorded consultations. Videos 1, 3 and 4 will be used to evaluate change of behaviour after training. Video-recorded consultations will be assessed by four independent and blinded (to the time-point of the video-recording) raters, using a pre-defined rating tool.

Participant level evaluation

Patient interviews

These interviews will be carried out within the four intervention practices. In-depth semi-structured interviews will be undertaken to explore patients' experiences of

the MOAC intervention. The purpose is to understand whether the advice and support offered is relevant to patients, whether they have implemented any of the recommendations, and how this has affected their perception of OA and its management.

For MOAC 1, individual in-depth interviews will be carried out in order to ascertain people's personal experiences and perspectives of their OA and the consultation. For MOAC 2, dependent on patient preference, either individual or group interviews will be held - bringing together three or four participants per group. The main reason for choosing the group interview is to avoid people feeling that they are being 'checked up on' with regards to implementing the advice and support offered in MOAC 2. The discussion within a group setting allows for the focus to be on the relevance and adoption of the intervention rather than on individuals.

The sampling frame for this study will be the MOSAICS consultation questionnaire described previously. In order to explore a range of accounts and behaviours, purposive sampling is required. Further variation will be achieved by selecting patients according to type of pain, age and gender. All interviews will be digitally-recorded and fully transcribed.

Two groups of patients will be purposively sampled:

1. Patients who have attended MOAC 1 will be interviewed soon after they have consulted their GP for their joint problem. An interview guide will provide a flexible framework for questioning, asking, for example; how do you feel the consultation went? Have you done anything differently? What information was given by the GP, and was it relevant/appropriate? How did you feel about being referred to a nurse? Approximately 15 people will be interviewed about their experiences of receiving MOAC 1.
2. Patients who have attended MOAC 2 will, dependent upon patient preference, either be interviewed individually or in small groups after completing up to four consultations with the nurse and returning their three-month consultation questionnaire. An interview guide will also be used and will focus questions around the expectations of the nurse consultation and what happened during the consultation, how the guidebook was used and how they felt about it, and overall what worked and what didn't. Approximately 15 people will be interviewed about their experience of receiving MOAC 2. These 15 people may not be the same as those interviewed about MOAC 1.

Evaluation of the training

Training evaluation questionnaires

Before training, to explore drivers for participation in the study, all GPs and practice nurses in participating practices will be invited to answer two open-ended questions that will focus on (i) their reasons for participating in the study and (ii) the perceived benefits (for clinical practice, for the primary care organisation, for their own role etc.). A baseline questionnaire assessment of knowledge about and beliefs and attitudes to OA and its management in primary care, will also be conducted. We will ask the GPs and practice nurses in the intervention practices who have received training to complete two further questionnaires (one month and six months post-training) to determine any change in knowledge, beliefs and attitudes following the training.

All GPs and practices nurses from the control practices will have the opportunity to attend the training at the end of the study. They will also be asked to complete the same evaluation questionnaire as the intervention practice staff at baseline (pre-training) and one month after the training (post-training).

Health economics evaluation

The economic evaluation will provide a preliminary analysis of the cost effectiveness (cost-utility) of the template and MOAC intervention compared with template alone, over a 12-month period.

Objectives: Study component 4

4a. To determine whether the training changes health care professionals' behaviour.
4b. To explore change in recording in medical records, and the uptake of core treatments for OA recommended by NICE in all participants consulting with joint pain (hip, knee, hand, foot) and the subgroup coded as OA by the general practitioner, following the MOAC intervention.
4c. To explore the cost-effectiveness of the template and MOAC intervention compared with the template alone.
4d. To explore patient experiences of their consultation for joint pain and whether the new intervention (MOAC) is acceptable and feasible in primary care (qualitative work).
4e. To examine and evaluate the way in which a new consultation for OA can be embedded in routine clinical practice in primary care (implementation).

Analysis

Analysis of the cluster RCT

Baseline characteristics will be compared between treatment arms and presented at the level of: (i) GP-Practice clusters, and (ii) Patient characteristics.

Baseline data for GP-Practice characteristics will include data on the stratification variables for randomisation – *i.e.*, practice list size and the number of GP practitioners, median index level of deprivation for the Practice, mean

age and gender profile of the Practice populations. Baseline data for patients will include data pertaining to participants' demographic characteristics, joint problem, management of their joint problem, general health and quality of life.

Balance of baseline characteristics is particularly important to establish for cluster trials given the higher level unit of randomisation and thus the potential for bias in the selection and recruitment uptake of patients. No formal statistical testing will be carried out for differences in baseline characteristics except for the GPAQ communication sub scale [49], as this is measure of the uptake of the model OA consultation by GPs.

Descriptive statistics on mean scores for numerical outcomes and frequency counts and percentages for categorical data will be presented for outcome measures from the consultation questionnaire, stratified by study group (intervention or control). A Linear Mixed Model will be used to analyse primary outcome data (SF-12 PCS). Statistical testing of clinical and process outcomes between study groups will be performed using regression methods (adjusting for age, gender, baseline SF12-PCS and corresponding baseline value, of the outcome being measured, as covariates at the individual-patient level and practice size as a covariate at the practice level). A 3-level mixed-model (linear- or generalised- as appropriate to numerical and categorical outcome data, respectively) will be fitted to test for the effect of the MOAC intervention from baseline across follow-up, taking into account clustering by practice (level 3) and patient (level 2) and repeat follow-up measures (level 1). P-values and 95% Confidence Intervals will be provided with estimates of effect size in the analysis of follow up data. The mixed model assumes that missing data is at least missing at random (MAR). We will examine effect estimates in relation to the indicated clinical marker of 0.3 of an effect size.

Complier Average Causal Effect Analysis (CACE)

A CACE analysis will be performed to provide an unbiased estimate of treatment effect for patients treated as per protocol specification (treatment administered as per protocol in the intervention arm is based on participants having seen the practice nurse MOAC 2 in the intervention practices).

Key secondary outcomes will be analysed including the Arthritis Self-Efficacy pain subscale [41] and the OMERACT/OARSI responder criteria [42], which combines measures of pain intensity (0 to 10 NRS) and function (subscale of the WOMAC) with global assessment of change to determine if participants are 'responders' to treatment.

Sub group analyses will be performed on the primary outcome (SF-12 PCS) as well as Arthritis Self-Efficacy pain sub-scale according to the following:

1. Age group (45 to 64 vs 65 and above);
2. Sex (male vs female);
3. Baseline SF12 – physical component score (cut off at median score);
4. Multi-site pain (less than 2 vs ≥2).

An interaction term (product of the subgroup variable and study group) will be included as an additional term in the regression models to evaluate the subgroup effect.

Sensitivity analyses will be conducted (this will be carried out on primary outcome [SF-12 PCS]), details of which can be found in Additional file 5.

Analysis of medical record data

We will split the medical records into three time periods: the 12 months before the template installation, the 6 months after installation but prior to randomisation, and the 12 months after start of the intervention. Eligible patients with OA or joint pain consultations in each time period (aged 45 years and over) will be identified. To assess the effect of the template, we will determine changes in routine recording of OA management pre- and post-template installation but prior to randomisation. This will include assessment of prescribing behaviours (paracetamol, topical NSAIDs, opioid, oral NSAIDs with or without a PPI, weight loss agents), investigation (use of relevant X-rays), and referral to selected specialities including exercise referral or physiotherapy, occupational therapy, weight loss programmes, orthopaedics, pain management, and rheumatology. We will determine the level of use of the template and the proportion of patients who have evidence of achievement of each Quality Indicator as recorded in the template in the first six months after installation. We will assess socio-demographic and clinical factors associated with achievement. Differences between the two clusters (template v. template plus intervention) on practice level outcomes will be assessed for the 12 months after the start of the intervention. The analysis will compare (i) between clusters on each of the Quality Indicator outcome measures adjusting for baseline achievement, and (ii) within clusters on change on each of the Quality Indicators from the six months pre-randomisation using multilevel modelling to adjust for clustering of patients within practices.

Analysis of HCP behaviour

General characteristics and baseline views of all GPs and nurses involved in the study will be determined. We will evaluate whether the knowledge, attitudes and beliefs, and reported practice of GPs and nurses from the intervention practices are (a) similar to the GPs and nurses from the control practices and (b) change as a result of the study training programme and involvement with OA patients over the time course of the study.

The video-recorded simulated patient consultations will be rated for the presence or absence of a number

of consultation behaviours (items), *e.g.*, did the GP elicit the patient's ideas or concerns about what the patient thinks is the matter, or did the GP tell the patient that the problem is due to OA? The overall rating for each GP for each consultation (the number of items rated as present) and the overall rating for each item at each time point (the number of GPs rated as having demonstrated that item) will be determined. The first value is a measure of the competency of a GP in delivering the model OA consultation (GP competency), and the second is a measure of the competency of all the GPs in delivering one element of the model OA consultation (item competency).

The following analyses will be undertaken:

1. The number of GPs who have increased GP competency at one month and six months post training.
2. The change in mean GP competency at one month and six months post training.
3. The change in item competency at one month and six months post training.

Analysis of qualitative data

Data will be entered into NVIVO 9 software to aid analysis. The constant comparative method [50] will be the primary analytical tool, but supplemented where appropriate by narrative analysis. The use of different data sources and methods allows for triangulation, and the development of the coding scheme. The qualitative team will carry out joint data analysis and interpretation of the Normalisation Process Theory (NPT) in relation to the data collected. The NPT will form the theoretical framework to order emerging themes and concepts across the intervention practices studied. The model focuses on the work that is required to get a new intervention integrated and workable. The key elements are: (i) the examination of how people as individuals and as a group make certain practices a reality; (2) what mechanisms promote or inhibit new practices; (3) the way in which practices are (re)produced by continuous investment by key HCPs so that it becomes part of clinical routines. By using the NPT, the new supported self-management strategy (the MOAC intervention and guidebook) will be studied in its totality, so that the findings can be more robust in terms of answering the question: does the new intervention work, under what circumstance, for whom and why?

Data analysis will take place in several stages. All interviews will then be coded, and coding schemes will be revised according to ongoing data analysis. Coded data will then be compared and emerging themes discussed. The quantitative findings will be embedded within the reporting of the analysis of the key quantitative results of the study.

Health economics analysis

The full Health Economics analysis plan is reported in Additional file 6.

Discussion

Despite the publication of National and International treatment recommendations, evidence suggests that there is a gap between the recommendations and what patients actually receive in the UK [11]. The NICE guidance (2008) highlights the possible therapeutic gains of positive self-management in primary care; however, there is as yet no evidence regarding the feasibility of implementing NICE core OA recommendations in primary care and the effect of this package on the course and impact of the condition [13]. To our knowledge, the MOSAICS study is the first to develop and evaluate a system for delivering these core messages in UK primary care.

Complex interventions are frequently employed in the NHS. Trials of complex interventions are of increasing importance because of the drive to provide the most cost effective healthcare; however, there are issues in describing, developing, documenting and reproducing complex interventions [23]. While such trials can pose a considerable challenge for researchers, approaches that incorporate both qualitative and quantitative evidence should lead to improved study design, implementation, evaluation and generalisability of results.

The MOSAICS study is a complex intervention and will use a cluster RCT design to test a novel intervention designed to increase the uptake of the core NICE OA guidelines. The primary aim is to evaluate a new model of supported self-management for OA, and evaluate the impact of this on practice level and HCP outcomes and in patients consulting with joint pain. Secondary aims are to: develop a training package for GPs and practice nurses; test the feasibility of recording the management of OA in consultations using 'quality markers' collected via a new consultation template; and describe the uptake of core NICE OA recommendations.

In line with the recommendations of the Medical Research Council (MRC) [51], this study includes a process evaluation component, which aims to explain any differences between what is expected and what actually happens in practice; and a health economic evaluation component, which will make the results of the evaluation more valuable to decision-makers.

Trial monitoring

The Research Centre's independent Data Monitoring Committee (DMC) will monitor the trial, and reports will be written in line with Arthritis Research UK recommendations (www.arthritisresearchuk.org). The trial will also be monitored by an independent trial steering committee (TSC) annually. This committee is made up of individuals

with expertise in musculoskeletal medicine, community rheumatology, biostatistics, nursing, community-based research, and health economics. The committee also includes a representative from Arthritis Care and an OA research user group. Both committees will be notified of any serious adverse events that may occur during the trial.

Patient and public involvement (PPI)

The Arthritis Research UK Primary Care Centre at Keele is committed to taking an explicit and systematic approach to involving patients and the public in research. For the MOSAICs study, the Research Users' group will work in collaboration with researchers on a wide range of tasks. These tasks will focus on aspects of research design, management and dissemination. Some examples of involvement include: development and design of the OA guidebook, advice on the content of the population survey, development of Quality Indicators for general practice consultations, involvement in developing training for HCPs and Steering Committee Membership.

Trial sponsor: Keele University

The sponsor will have no role in the design and analysis of the data.

Additional files

Additional file 1: Resources provided.
Additional file 2: Consent form.
Additional file 3: The MOAC intervention.
Additional file 4: List of Read codes.
Additional file 5: Sensitivity analyses.
Additional file 6: Health economics analysis plan.

Competing interests

The authors declare that they have no competing interest.

Authors' contribution

All authors participated in the design of the study and drafting the manuscript. All authors read and approved the final manuscript.

Authors' information

Chief Investigator: Professor Krysia Dziedzic.
Population survey: Principal Investigators: Professor Krysia Dziedzic and Dr. Martyn Lewis Consultation data and template: Principal Investigators: Dr. John Edwards and Dr. Kelvin Jordan.
Intervention & evaluation of training:
Principal Investigator (MOAC 1): Dr. Mark Porcheret.
Principal Investigators (MOAC 2): Dr. Emma Healey and Professor Chris Main.
Principal Investigator (MOAC 3): Professor Krysia Dziedzic.
Qualitative evaluation:
Principal Investigators: Dr Clare Jinks and Professor Pauline Ong.
Trial coordinator: Kris Clarkson.
Trial Statistician: Dr. Martyn Lewis with support from Ebenezer Afolabi.
Health Economist: Dr. Sue Jowett with support from Raymond Oppong.
User involvement: Dr. Clare Jinks.

Acknowledgements

This paper presents independent research commissioned by the National Institute for Health Research (NIHR) Programme Grant (RP-PG-0407-10386). The views expressed in this paper are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. This research was also funded by the Arthritis Research UK Centre in Primary Care grant (Grant Number 18139). EH and CJ are part-funded by the National Institute for Health Research (NIHR) Collaborations for Leadership in Applied Health Research and Care West Midlands. JJE is supported by an In-Practice Fellowship from the National Institute for Health Research (NIHR) and AF is supported by an NIHR Doctoral, Clinical Academic Training Fellowship. The authors would like to thank the OA Research Users' Group and the network, nursing, health informatics and administrative staff at Keele University's Arthritis Research UK Primary Care Centre for all their support and assistance with this study. The authors would like to give special thanks to all of the staff at the participating general practices and the GP facilitators Dr. Marwyn Ryles, Dr. Tayo Shoran and Dr. Beth Hanson, who provided support to the general practices involved in the study. Thanks go to Professor Robbie Foy, Professor Martin Eccles, Professor Susan Michie, Janet Grime and Vince Cooper for their valued thoughts and suggestions on the developmental work for this trial. The authors would also like to thank all members of the Trial Steering Committee and Data Monitoring Committee for their valuable input in study design and conduct.

Author details

[1]Arthritis Research UK Primary Care Centre, Keele University, Keele, Staffordshire ST5 5BG, UK. [2]The School of Nursing, Midwifery and Social Work, The University of Manchester, Manchester M13 9PL, UK. [3]The Haywood Hospital, Staffordshire Rheumatology Centre, High Lane, Burslem, Stoke-On-Trent, Staffordshire ST6 7AG, UK. [4]School of Health and Population Sciences, College of Medical and Dental Sciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK. [5]Nottingham Clinical Trials Unit, Nottingham Health Science Partners, C Floor, South Block, Queen's Medical Centre, Nottingham NG7 2UH, UK.

Received: 17 April 2014 Accepted: 8 July 2014
Published online: 27 August 2014

References

1. Jinks C, Ong BN, Richardson J: **A mixed methods study to investigate needs assessment for knee pain and disability: population and individual perspectives.** *BMC Musculoskelet Disord* 2007, **8**:59.
2. Hill S, Dziedzic K, Thomas E, Baker SR, Croft P: **The illness perceptions associated with health and behavioural outcomes in people with musculoskeletal hand problems: findings from the North Staffordshire OA Project (NorStOP).** *Rheumatology* 2007, **46**:944–951.
3. Altman RD, Hochberg MC, Moskowitz RW, Schnitzer TJ: **Recommendations for the medical management of OA of the hip and knee: 2000 update.** *Arthritis Rheum* 2000, **43**:1905–1915.
4. Jordan KM, Arden NK, Doherty M, Bannwarth B, Bijlsma JWJ, Dieppe P, Gunther K, Hauselmann H, Herrero-Beaumont G, Kaklamanis P, Lohmander S, Leeb B, Lequesne M, Mazieres B, Martin-Mola E, Pavelka K, Pendleton A, Punzi L, Serni U, Swoboda B, Verbruggen G, Zimmerman-Gorska I, Dougados M, Standing Committee for International Clinical Studies Including Therapeutic Trials ESCISIT: **EULAR Recommendations 2003: an evidence based approach to the management of knee OA: Report of a Task Force of the Standing Committee for International Clinical Studies Including Therapeutic Trials (ESCISIT).** *Ann Rheum Dis* 2003, **62**:1145–1155.
5. Zhang W, Moskowitz RW, Nuki G, Abramson S, Altman RD, Arden N, Bierma-Zeinstra S, Brandt KD, Croft P, Doherty M, Dougados MD, Hochberg M, Hunter DJ, Kwoh K, Lohmander LS, Tugwell P: **OARSI recommendations for the management of hip and knee OA, Part I: Critical appraisal of existing treatment guidelines and systematic review of current research evidence.** *OA Cartilage* 2007, **15**:981–1000.
6. Zhang W, Doherty M, Leeb BF, Arden N, Bijlsma J, Dincer F, Dziedzic K, Häuselmann HJ, Herrero-Beaumont G, Kaklamanis P, Lohmander S, Maheu E, Martin-Mola E, Pavelka K, Punzi L, Reiter S, Sautner J, Smolen J, Verbruggen G, Zimmermann-Górska I: **EULAR evidence based recommendations for the management of hand OA - report of a task force of the EULAR Standing Committee for International Clinical Studies Including Therapeutics (ESCISIT).** *Ann Rheum Dis* 2007, **66**:377–388.
7. Zhang W, Moskowitz RW, Nuki G, Abramson S, Altman RD, Arden N, Bierma-Zeinstra S, Brandt KD, Croft P, Doherty M, Dougados M, Hochberg M, Hunter DJ, Kwoh K, Lohmander LS, Tugwell P: **OARSI recommendations for the**

management of hip and knee OA, Part II: OARSI evidence-based, expert consensus guidelines. *OA Cartilage* 2008, 16:137–162.
8. Zhang W, Doherty M, Leeb BF, Alekseeva L, Arden N, Bijlsma J, Dincer F, Dziedzic K, Hauselmann HJ, Kaklamanis P, Kloppenburg M, Lohmander LS, Maheu E, Martin-Mola E, Pavelka K, Punzi L, Reiter S, Smolen J, Verbruggen G, Watt I, Zimmermann-Gorska I: **EULAR evidence based recommendations for the diagnosis of hand OA - report of a taskforce of the EULAR Standing Committee for International Clinical Studies Including Therapeutics (ESCISIT).** *Ann Rheum Dis* 2009, **68**:8–17.
9. Zhang W, Doherty M, Peat G, Bierma-Zeinstra MA, Arden NK, Bresnihan B, Herrero-Beaumont G, Kirschner S, Leeb BF, Lohmander LS, Mazières B, Pavelka K, Punzi L, So AK, Tuncer T, Watt I, Bijlsma JW: **EULAR evidence based recommendations for the diagnosis of knee OA.** *Ann Rheum Dis* 2010, **69**:483–489. doi:10.1136/ard.2009.113100.
10. Conaghan PG, Dickson J, Grant RL: **Care and management of OA in adults: summary of NICE guidance.** *BMJ* 2008, **336**:502–503.
11. Porcheret M, Jordan K, Jinks C, with the Primary Care Rheumatology Society: **Primary care treatment of knee pain a survey in older adults.** *Rheumatol* 2007, **46**:1694–1700.
12. Steel N, Bachmann M, Maisey S, Shekelle P, Breeze E, Marmot M, Melzer D: **Self reported receipt of care consistent with 32 Quality Indicators: national population survey of adults aged 50 or more in England.** *BMJ* 2008, **337**:a957.
13. NICE: **OA: The care and management of adults with OA.** In *National Institute of Health and Clinical Excellence*. 2008. http://www.nice.org.uk/guidance/CG59.
14. Finney A, Porcheret M, Grime J, Jordan KP, Handy J, Healey EL, Ryan S, Jester R, Dziedzic KS: **Defining the content of an opportunistic osteoarthritis consultation with primary health care professionals: A dephi consensus study.** *Arthritis Care Res* 2013, **65**:962–8. doi:10.1002/acr.21917.
15. Porcheret M, Grime J, Main C, Dziedzic K: **Developing a model osteoarthritis consultation: a Delphi consensus exercise.** *BMC Musculoskelet Disord* 2013, **14**:25.
16. Department of Health: *The NHS improvement plan: putting people at the heart of public services*. London: Department of Health; 2004.
17. Mann C, Gooberman-Hill R: **Health Care Provision for Osteoarthritis Concordance Between What Patients Would Like and What Health Professionals Think They Should Have.** *Arthritis Care Res* 2011, **63**:963–972. doi:10.1002/acr.20459.
18. Kralik D, Koch T, Price K, Howard N: **Chronic illness self-management: taking action to create order.** *J Clin Nurs* 2004, **13**:259–267.
19. Thorne S, Paterson BL, Russell C: **The structure of everyday self-care decision making.** *Qual Health Res* 2003, **13**:1337–1352.
20. Townsend A, Wyke S, Hunt K: **Self-managing and managing self: practical and moral dilemmas in accounts of living with chronic illness.** *Chronic Illn* 2006, **2**:185–195.
21. Kendall E, Rogers A: **Extinguishing the social?: state sponsored self-care policy and the Chronic Disease Self-management Programme.** *Disability Soc* 2007, **22**:129–143.
22. Taylor D, Bury M: **Chronic illness, expert patients and care transition.** *Social Health Illness* 2007, **29**:27–45.
23. Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, Tyrer P: **Framework for design and evaluation of complex interventions to improve health.** *BMJ* 2010, **340**:c869. doi:10.1136/bmj.c869.
24. Kennedy AP, Rogers AE, Bower P: **Support for self care for patients with chronic disease.** *BMJ* 2007, **335**:968–970.
25. Kennedy AP, Nelson E, Reeves D, Richardson G, Roberts C, Robinson A, Rogers AE, Sculpher M, Thompson DG, the North-West Regional Gastrointestinal Research Group: **A randomised controlled trial to assess the effectiveness and cost of a patient orientated self-management approach to chronic inflammatory bowel disease.** *Gut* 2004, **53**:1639–1645.
26. Robinson A, Lee V, Kennedy A, Middleton L, Rogers A, Thompson DG, Reeves D: **A randomised controlled trial of self-help interventions in patients with a primary care diagnosis of irritable bowel syndrome.** *Gut* 2006, **55**:643–648.
27. Kennedy A, Bower P, Reeves D, Blakeman T, Bowen R, Chew-Graham C, Eden M, Fullwood C, Gaffney H, Gardner C, Lee V, Morris R, Protheroe J, Richardson G, Sanders C, Swallow A, Thompson D, Rogers A: **Implementation of self management support for long term conditions in routine primary care settings: cluster randomised controlled trial.** *BMJ* 2013, **346**:f2882. doi:http://dx.doi.org/10.1136/bmj.f2882.
28. Michie S, Johnston M, Abraham C, Lawton R, Parker D, Walker A: **Making psychological theory useful for implementing evidence based practice: a consensus approach.** *Qual Saf Health Care* 2005, **14**:26–33. doi:10.1136/qshc.2004.011155.
29. Michie S, Johnston M, Francis J, Hardeman W, Eccles M: **From theory to Intervention: Mapping Theoretically Derived Behavioural Determinants to Behaviour Change Techniques.** *Applied Psychol* 2008, **57**:66–680.
30. Michie S, van Stralen MM, West R: **The behaviour change wheel: A new method for characterising and designing behaviour change interventions.** *Implementation Sci* 2011, **6**:42. doi:10.1186/1748-5908-6-42.
31. Grol R: **Beliefs and evidence in changing clinical practice.** *BMJ* 1997, **315**:418–421.
32. Silverman J, Kurt S, Draper J: *Skills for Communicating with Patients*. 2nd edition. Oxford: Radcliffe Publishing; 2005.
33. May T: *Social Research: Issues, methods and process*. Buckingham: Open University Press; 1993.
34. May C, Finch T: **Implementing, embedding and integrating practices: an outline of Normalisation Process Theory.** *Sociol* 2009, **43**:535–554.
35. Chan A-W, Tetzlaff JM, Altman DG, Laupacis A, Gøtzsche PC, Krleža-Jerić K, Hróbjartsson A, Mann H, Dickersin K, Berlin J, Doré C, Parulekar W, Summerskill W, Groves T, Schulz K, Sox H, Rockhold FW, Rennie D, Moher D: **SPIRIT: Statement: Defining standard protocol items for clinical trials.** *Ann Intern Med* 2013, **2013**(158):200–207.
36. Chan A-W, Tetzlaff JM, Gøtzsche PC, Altman DG, Mann H, Berlin J, Dickersin K, Hróbjartsson A, Schulz KF, Parulekar WR, Krleža-Jerić K, Laupacis A, Moher D: **SPIRIT 2013b Explanation and Elaboration: Guidance for protocols of clinical trials.** *BMJ* 2013, **346**:e7586.
37. Dziedzic KS, Healey EL, Main CJ: **Implementing the NICE osteoarthritis guidelines in primary care: a role for practice nurses.** *Musculoskeletal Care* 2013, **11**:1. doi:10.1002/msc.1040.
38. Grime J, Dudley B: **Developing written information on OA for patients: facilitating user involvement by exposure to qualitative research.** *Health Expect* 2011. doi:10.1111/j.1369-7625.2011.00741.
39. Norman GR, Sloan JA, Wyrwich KW: **Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation.** *Med Care* 2003, **41**:582–92.
40. Ware JE, Kosinski M, Keller SD: **A 12-item short form health survey (SF-12): Construction and preliminary tests of reliability and validity.** *Med Care* 1996, **32**:220–33.
41. Lorig K, Chastain R, Ung E, Shoor S, Holman HR: **Development and evaluation of a scale to measure perceived self-efficacy in people with arthritis.** *Arthritis Rheum* 1989, **32**:37–44.
42. Pham T, van der Heijde D, Altman RD, Anderson JJ, Bellamy N, Hochberg M, Simon L, Strand V, Woodworth T, Dougados M: **OMERACT-OARSI initiative: Osteoarthritis Research Society International set of responder criteria for osteoarthritis clinical trials revisited.** *Osteoarthritis Cartilage* 2004, **12**:389–399.
43. Angrist JD, Imbens GW, Rubin DB: **Identification of causal effects using instrumental variables.** *J Am Stat Assoc* 1996, **91**:444–455.
44. Spitzer RL, Kroenke K, Williams JB, Löwe B: **A brief measure for assessing generalized anxiety disorder: the GAD-7.** *Arch Intern Med* 2006, **166**:1092–1097.
45. Group EQ: **EuroQol: a new facility for the measurement of health-related quality of life.** *Health Policy* 1990, **16**:199–208.
46. Matthews CE, Ainsworth BE, Hanby C, Pate RR, Addy C, Freedson PS, Jones DA, Macera CA: **Development and testing of a short physical activity recall questionnaire.** *Med Sci Sports Exerc* 2005, **37**:986–994.
47. Jinks C, Jordan K, Ong BN, Croft P: **A brief screening tool for knee pain in primary care (KNEST) 2. Results from a survey in the general population aged 50 and over.** *Rheumatology* 2004, **43**:55–61.
48. Edwards JJ, Khanna M, Jordan KP, Jordan JL, Bedson J, Dziedzic KS: **Quality indicators for the primary care of osteoarthritis: a systematic review.** *Ann Rheum Dis* 2013. Published On-line First: 27 November 2013. doi:10.1136/annrheumdis-2013-203913.
49. Mead N, Bower P, Roland M: **The General Practice Assessment Questionnaire (GPAQ) – Development and psychometric characteristics.** *BMC Fam Pract* 2008, **9**:13. doi:10.1186/1471-2296-9-13.
50. Charmaz K: *Constructing grounded theory: A Practical Guide through Qualitative Analysis*. London: Sage Publications Ltd; 2006.
51. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M: **Developing and evaluating Complex interventions: new guidance.** *MRC* 2008. http://www.mrc.ac.uk/documents/pdf/complex-interventions-guidance/.

52. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW: Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol* 1988, 15:1833–1840.
53. Meenan RF, Mason JH, Anderson JJ, Guccione AA, Kazis LE: AIMS2: The content and properties of a revised and expanded Arthritis Impact Measurement Scales Health Status Questionnaire. *Arthritis Rheum* 1992, 35:1–10.
54. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, Pratt M, Ekelund U, Yngve A, Sallis JF, Oja P: International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc* 2003, 35:1381–1395.
55. Washburn RA, Smith K, Jette AM, Janney CA: The physical activity scale for the elderly (PASE): development and evaluation. *J Clin Epi* 1993, 46:153–162.
56. van der Windt DA, van der Heijden GJ, de Winter AF, Koes BW, Deville W, Bouter L: The responsiveness of the Shoulder Disability Questionnaire. *Ann Rheum Dis* 1998, 57:82–87.
57. Kroenke K, Strine TW, Spitzer RL, Williams JBW, Berry JT, Mokdad AH: The PHQ-8 as a measure of current depression in the general population. *J Affective Disorders* 2009, 114:163–173.
58. Brazier JE, Roberts J: The Estimation of a Preference-Based Measure of Health from the SF-12. *Med Care* 2004, 42:851–859.
59. Al-Janabi H, Flynn T, Coast J: Development of a self-report measure of capability wellbeing for adults: the ICECAP-A. *Qual Life Res* 2012, 21:167–176.
60. Østerås N, Garratt A, Grotle M, Natvig, Kjeken I, Kvien TK, Hagen KB: Patient-reported quality of care for osteoarthritis: development and testing of the osteoarthritis quality indicator questionnaire. *Arthritis Care Res* 2013, 65:1043–1051.
61. Howie J, Heaney DJ, Maxwell M, Walker JJ: A comparison of a Patient Enablement Instrument (PEI) against two established satisfaction scales as an outcome measure of primary care consultations. *Fam Pract* 1998, 15:165–171.

doi:10.1186/s13012-014-0095-y
Cite this article as: Dziedzic *et al.*: Implementing the NICE osteoarthritis guidelines: a mixed methods study and cluster randomised trial of a model osteoarthritis consultation in primary care - the Management of OsteoArthritis In Consultations (MOSAICS) study protocol. *Implementation Science* 2014 9:95.

Appendix 1.2 MOSAICS ethical approval letter

NHS

National Research Ethics Service

North West 1 Research Ethics Committee – Cheshire
Research Ethics Office
Barlow House
3rd Floor
4 Minshull Street
Manchester
M1 3DZ

Telephone: 0161 625 7821

18 November 2010

Professor Krysia Dziedzic
Arthritis Research UK Professor of Musculoskeletal Therapies
Arthritis Research UK Primary Care Centre
Primary Care Sciences
Keele University
Staffordshire
ST5 5BG

Dear Professor Dziedzic

Full title of study: **Management of Osteoarthritis in Consultations Study: the development of a complex intervention in primary care (MOSAICS)**
REC reference number: **10/H1017/76**

Thank you for your letters of 15th and 16th November 2010. I can confirm the REC has received the documents listed below as evidence of compliance with the approval conditions detailed in our letter dated 26 October 2010. Please note these documents are for information only and have not been reviewed by the committee.

I can confirm that the additional conditions have now been met.

Documents received

The documents received were as follows:

Document	*Version*	*Date*
Covering Letter		16 November 2010
Covering Letter		15 November 2010
Participant Information Sheet: Appendix 4.b.3.2 MCQ Information Leaflet (Questionnaires and Interviews)	2	26 October 2010
Participant Consent Form: Appendix 4.e.9 Manager and administrator Consent Form	2	16 November 2010
Participant Information Sheet: Appendix 4.d.13 Patient information leaflet MOAC 3	2	26 October 2010
Participant Information Sheet: Appendix 4.d.7 Patient information leaflet Group interview MOAC 2	2	26 October 2010
Participant Consent Form: Appendix 4.e.1.2 - Nurse Consent Form Observation MOAC 2	2	16 November 2010
Participant Consent Form: Appendix 4.e.5 HCP Interview Consent Form	2	16 November 2010
Questionnaire: Appendix 2.3 Population Survey Questionnaire	2	15 November 2010

You should ensure that the sponsor has a copy of the final documentation for the study. It is the sponsor's responsibility to ensure that the documentation is made available to R&D offices at all participating sites.

10/H1017/76	**Please quote this number on all correspondence**

Yours sincerely

Miss Shehnaz Ishaq
Committee Co-ordinator

Email: Shehnaz.ishaq@northwest.nhs.uk

Copy to:

Dr Mark Porcheret
Arthritis Research UK Primary Care Centre
Primary Care Sciences
Keele University
Keele
ST5 5BG

Ms June Handy
Arthritis Research UK Primary Care Centre
Primary Care Sciences
Keele University
Keele
ST5 5BG

Rhian Hughes
Arthritis Research UK Primary Care Centre
Primary Care Sciences
Keele University
Keele
ST5 5BG

Ms Nemonie Marriott
Stoke-on-Trent PCT
London House
4th Floor
Hide Street
Stoke-on-Trent, ST4 1NF

Appendix 1.3 MOSAICS population Survey joint pain

Section B. Pain in different parts of the body

Part 1 - Specific joint pains and problems in the last year

We are interested to know how many people suffer from specific joint pains and problems. Please answer each of the following questions about your hips, knees, hands and feet, **even if you do not suffer from any problems.**

1. Have you had any pain in the last year in or **around the HIP**?
   (Please place a cross in one box only)

   Yes..... ☐ No..... ☐

2. Have you had any pain in the last year in and **around the KNEE?**
   (Please place a cross in one box only)

   Yes..... ☐ No..... ☐

3. Have you had any pain in the last year in or **around the HAND**?
   (Please place a cross in one box only)

   Yes..... ☐ No..... ☐

4. Have you had any pain in the last year in or **around the FOOT**?
   (Please place a cross in one box only)

   Yes..... ☐ No..... ☐

Appendix 1.4 MOSAICS Population Survey joint pain intensity

Part 2 – Pain in different parts of the body in the last month

This is about pain in different parts of your body in the last month. Please complete each of the following questions **even if you have not suffered pain in any of these areas in the last month**.

1. In the **past month**, **on average**, **how intense were each of these pains** rated on a 0-10 scale where 0 is "no pain" and 10 is "pain as bad as could be"? (That is, your usual pain at times you were experiencing pain.)

For each pain, please put a cross in one box. For pains that do not apply to you please put a cross in box 0 to indicate No Pain.

		No Pain										Pain as bad as could be
		0	1	2	3	4	5	6	7	8	9	10
a.	Hip pain	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
b.	Knee pain	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
c.	Hand pain	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
d.	Foot pain	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Appendix 1.5 Systematic Review Stage A Article Selection Form

First Author	Journal	Year	Do participants fit?(Y/N)	Assess Physical activity instrument? (Y/N)	Include in the Full paper review? (Y/N)	Comments

Appendix 1.6 Systematic Review Stage A Data Extraction Form

Authors	Year	Journal	Include?	Study Design	Subjects	Physical activity measurement used?	References of interest for Stage 2?	Article for Stage 2?	Instrument Name For Stage 2?	Comments

Appendix 1.7 Systematic Review Stage B Article Selection Form

First Author	Journal	Year	Do participants fit?(Y/N)	Assess Physical activity instrument? (Y/N)	Include in the Full paper review? (Y/N)	Comments

Appendix 1.8 Systematic Review Stage B Modified from Terwee et al. (2010)

QAPAQ

Data Extraction form for the Qualitative attributes and the measurement properties of physical activity questionnaires (QAPAQ).

A data extraction form should be used for each study, this is a data extraction not quality assessment, reviewer only needs to report what is in the article, all critical analysis of article should go in COSMIN. There will be many times where there are items in the form that the article does not cover, please leave these blank. In cases where a single article assess many different instruments please use a new data extraction form for each instrument.

First Author:

Year:

Journal:

Instrument Name:

Measurement Properties	Appraisal of Questionnaire's Measurement Properties
General Issues	
Study's population? (setting, health condition, country)	
Population age	
Population sex	
Sample size (n= total sample size)	
Version of questionnaire used	
Time intervals of administration	
Mode of administration	
Other instruments used	

Reliability	
Measurement Error	
Reliability intra class correlation	
Validity	
	In this section only fill out section what is covered in the article by the authors, don't answer the questions yourself. If sections are not included in article please leave blank.
Face Validity	
Is the question asked to allow for an accurate answer?	
Does the combination of items into a score make sense?	
Are all items comprehensive?	
Content Validity	
Is the instrument comprehensive?	
Are frequency, duration and intensity addressed?	
Is there enough details in the questionnaire for subjects to answer appropriately on frequency, intensity and duration in the recall period?	
Is justification given for choices made in the instrument?	
Are there any important questions missing?	
Are there any unsuitable questions asked?	
Are the scales used too coarse or too fine?	

Is the weighing of the items scoring appropriate for a total score?	
Floor Ceiling Effect	
Number of responders achieving lowest or highest score?	
Construct Validity	
Responsiveness	

Appendix 1.9 Systematic Review Stage B Modified COSMIN Quality Assessment Form

COSMIN checklist with 4-point scale

Reviewer name:	
Instrument:	
Study reference (first author and year):	

Evaluated measurement properties in the article	**Tick**	
Internal consistency		Box A
Reliability		Box B
Measurement error		Box C
Content validity		Box D
Structural validity		Box E
Hypotheses testing		Box F
Cross-cultural validity		Box G
Criterion validity		Box H
Responsiveness		Box I

Note: only complete boxes where the measurement property has been assessed. Do not complete all boxes if not necessary.
In the boxes some items may not apply to the study, you can make a NA anywhere in the row where item does not apply to the study. – You do not need to complete all items in the boxes if those items are not applicable.

Internal Consistency: Box A

Item	Excellent	tick	Good	tick	Fair	tick	Poor	tick
Was the percentage of missing items given?	Percentage of missing items described		Percentage of missing items NOT described		-		-	
Was there a description of how missing items were handled?	Described how missing items were handled		Not described but it can be deduced how missing items were handled		Not clear how missing items were handled			
Was the sample size included in the internal consistency analysis adequate?	Adequate sample size (≥100)		Good sample size (50-99)		Moderate sample size (30-49)		Small sample size (<30)	
Was the unidimensionality of the scale checked? i.e. was factor analysis	Factor analysis performed in the study population		Authors refer to another study in which factor analysis was performed in a similar study population		Authors refer to another study in which factor analysis was performed, but not in a similar study population		Factor analysis NOT performed and no reference to another study	
Was the sample size included in the unidimensionality analysis adequate?	7* #items and ≥100		5* #items and ≥100 OR 6-7* #items but <100		5* #items but <100		<5* #items	
Was an internal consistency statistic calculated for each (unidimensional) (sub) scale separately?	Internal consistency statistic calculated for each subscale separately		-		-		Internal consistency statistic NOT calculated for each subscale separately	
Were there any important flaws in the design or	No other important methodological flaws in the design or execution of the study				Other minor methodological flaws in		Other important methodological	

methods of the study?					the design or execution of the study		flaws in the design or execution of the study	
for Classical Test Theory (CTT), continuous scores: Was Cronbach's alpha calculated?	Cronbach's alpha calculated		-		Only item-total correlations calculated		No Cronbach's alpha and no item-total correlations calculated	
for CTT, dichotomous scores: Was Cronbach's alpha or KR-20 calculated?	Cronbach's alpha or KR-20 calculated				Only item-total correlations calculated		No Cronbach's alpha or KR-20 and no item-total correlations calculated	

Reliability: Box B

Item	**Excellent**	**tick**	**Good**	**tick**	**Fair**	**tick**	**Poor**	**tick**
Was the percentage of missing items given?	Percentage of missing items described		Percentage of missing items NOT described		-		-	
Was there a description of how missing items were handled?	Described how missing items were handled		Not described but it can be deduced how missing items were handled		Not clear how missing items were handled		-	
Was the sample size included in the analysis adequate?	Adequate sample size (≥100)		Good sample size (50-99)		Moderate sample size (30-49)		Small sample size (<30)	
Were at least two measurements available?	At least two measurements		-		-		Only one measurement	
Were the administrations independent?	Independent measurements		Assumable that the measurements were independent		Doubtful whether the measurements were independent		measurements NOT independent	
Was the time interval stated?	Time interval stated		-		Time interval NOT stated		-	
Were patients stable in the interim period on the construct to be measured?	Patients were stable (evidence provided)		Assumable that patients were stable		Unclear if patients were stable		Patients were NOT stable	

Was the time interval appropriate?	Time interval appropriate		-		Doubtful whether time interval was appropriate		Time interval NOT appropriate	
Were the test conditions similar for both measurements? e.g. type of administration, environment, instructions	Test conditions were similar (evidence provided)		Assumable that test conditions were similar		Unclear if test conditions were similar		Test conditions were NOT similar	
Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		-		Other minor methodological flaws in the design or execution of the study		Other important methodological flaws in the design or execution of the study	
For continuous scores: Was an intraclass correlation coefficient (ICC) calculated?	ICC calculated and model or formula of the ICC is described		ICC calculated but model or formula of the ICC not described or not optimal. Pearson or Spearman correlation coefficient calculated with evidence provided that no systematic change has occurred		Pearson or Spearman correlation coefficient calculated WITHOUT evidence provided that no systematic change has occurred or WITH evidence that systematic change has occurred		No ICC or Pearson or Spearman correlations calculated	
For dichotomous/nominal/ordinal scores: Was kappa calculated?	Kappa calculated		-		-		Only percentage agreement calculated	
For ordinal scores: Was a weighted kappa calculated?	Weighted Kappa calculated		-		Unweighted Kappa calculated		Only percentage agreement calculated	
for ordinal scores: Was the weighting scheme described? e.g. linear, quadratic	Weighting scheme described		Weighting scheme NOT described		-		-	

Measurement error: Box C

Item	Excellent	tick	Good	tick	Fair	tick	Poor	tick
Was the percentage of missing items given?	Percentage of missing items described		Percentage of missing items NOT described		-		-	
Was there a description of how missing items were handled?	Described how missing items were handled		Not described but it can be deduced how missing items were handled		Not clear how missing items were handled		-	
Was the sample size included in the analysis adequate?	Adequate sample size (≥100)		Good sample size (50-99)		Moderate sample size (30-49)		Small sample size (<30)	
Were at least two measurements available?	At least two measurements		-		-		Only one measurement	
Were the administrations independent?	Independent measurements		Assumable that the measurements were independent		Doubtful whether the measurements were independent		measurements NOT independent	
Was the time interval stated?	Time interval stated		-		Time interval NOT stated		-	
Were patients stable in the interim period on the construct to be measured?	Patients were stable (evidence provided)		Assumable that patients were stable		Unclear if patients were stable		Patients were NOT stable	
Was the time interval appropriate?	Time interval appropriate		-		Doubtful whether time interval was appropriate		Time interval NOT appropriate	
Were the test conditions similar for both measurements? e.g. type of administration, environment, instructions	Test conditions were similar (evidence provided)		Assumable that test conditions were similar		Unclear if test conditions were similar		Test conditions were NOT similar	
Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		-		Other minor methodological flaws in the design or execution of the study		Other important methodological flaws in the design or execution of the study	

for CTT: Was the Standard Error of Measurement (SEM), Smallest Detectable Change (SDC) or Limits of Agreement (LoA) calculated?	SEM, SDC, or LoA calculated		Possible to calculate LoA from the data presented		-		SEM calculated based on Cronbach's alpha, or on SD from another population	

Content validity: Box D

Item	**Excellent**	**tick**	**Good**	**tick**	**Fair**	**tick**	**Poor**	**tick**
Was there an assessment of whether all items refer to relevant aspects of the construct to be measured?	Assessed if all items refer to relevant aspects of the construct to be measured		-		Aspects of the construct to be measured poorly described AND this was not taken into consideration		NOT assessed if all items refer to relevant aspects of the construct to be measured	
Was there an assessment of whether all items are relevant for the study population? (e.g. age, gender, disease characteristics, country, setting)	Assessed if all items are relevant for the study population in adequate sample size (≥10)		Assessed if all items are relevant for the study population in moderate sample size (5-9)		Assessed if all items are relevant for the study population in small sample size (<5)		NOT assessed if all items are relevant for the study population OR target population not involved	
Was there an assessment of whether all items are relevant for the purpose of the measurement instrument? (discriminative, evaluative, and/or predictive)	Assessed if all items are relevant for the purpose of the application		Purpose of the instrument was not described but assumed		NOT assessed if all items are relevant for the purpose of the application		-	
Was there an assessment of whether all items together comprehensively reflect the construct to be measured?	Assessed if all items together comprehensively reflect the construct to be measured		-		No theoretical foundation of the construct and this was not taken into consideration		NOT assessed if all items together comprehen-sively reflect the construct to be measured	

Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		-		Other minor methodological flaws in the design or execution of the study		Other important methodological flaws in the design or execution of the study	

Structural validity: Box E

Item	**Excellent**	**tick**	**Good**	**tick**	**Fair**	**tick**	**Poor**	**tick**
Was the percentage of missing items given?	Percentage of missing items described		Percentage of missing items NOT described		-		-	
Was there a description of how missing items were handled?	Described how missing items were handled		Not described but it can be deduced how missing items were handled		Not clear how missing items were handled		-	
Was the sample size included in the analysis adequate?	7* #items and ≥100		5* #items and ≥100 OR 5-7* #items but <100		5* #items but <100		<5* #items	
Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		-		Other minor methodological flaws in the design or execution of the study (e.g. rotation method not described)		Other important methodological flaws in the design or execution of the study (e.g. inappropriate rotation method)	
for CTT: Was exploratory or confirmatory factor analysis performed?	Exploratory or confirmatory factor analysis performed and type of factor analysis appropriate in view of existing information		Exploratory factor analysis performed while confirmatory would have been more appropriate		-		No exploratory or confirmatory factor analysis performed	

Construct validity: Box F

Item	**Excellent**	**tick**	**Good**	**tick**	**Fair**	**tick**	**Poor**	**tick**
Was the percentage of missing items given?	Percentage of missing items described		Percentage of missing items NOT described		-		-	
Was there a description of how missing items were handled?	Described how missing items were handled		Not described but it can be deduced how missing items were handled		Not clear how missing items were handled		-	

Was the sample size included in the analysis adequate?	Adequate sample size (≥100)		Good sample size (50-99)		Moderate sample size (30-49)		Small sample size (<30)	
Were hypotheses regarding correlations or mean differences formulated a priori (i.e. before data collection)?	Multiple hypotheses formulated a priori		Minimal number of hypotheses formulate a priori		Hypotheses vague or not formulated but possible to deduce what was expected		Unclear what was expected	
Was the expected direction of correlations or mean differences included in the hypotheses?	Expected direction of the correlations or differences stated		Expected direction of the correlations or differences NOT stated		-		-	
Was the expected absolute or relative magnitude of correlations or mean differences included in the hypotheses?	Expected magnitude of the correlations or differences stated		Expected magnitude of the correlations or differences NOT stated					
for convergent validity: Was an adequate description provided of the comparator instrument(s)?	Adequate description of the constructs measured by the comparator instrument(s)		Adequate description of most of the constructs measured by the comparator instrument(s)		Poor description of the constructs measured by the comparator instrument(s)		NO description of the constructs measured by the comparator instrument(s)	
for convergent validity: Were the measurement properties of the comparator instrument(s) adequately described?	Adequate measurement properties of the comparator instrument(s) in a population similar to the study population		Adequate measurement properties of the comparator instrument(s) but not sure if these apply to the study population		Some information on measurement properties (or a reference to a study on measurement properties) of the comparator instrument(s) in any study population		No information on the measurement properties of the comparator instrument(s)	
Were there any important flaws in the	No other important methodological flaws in the		-		Other minor methodological flaws in the design or execution		Other important methodological flaws in the design	

design or methods of the study?	design or execution of the study				of the study (e.g. only data presented on a comparison with an instrument that measures another construct)		or execution of the study	
Were design and statistical methods adequate for the hypotheses to be tested?	Statistical methods applied appropriate		Assumable that statistical methods were appropriate, e.g. Pearson correlations applied, but distribution of scores or mean (SD) not presented		Statistical methods applied NOT optimal		Statistical methods applied NOT appropriate	

Cross cultural validity: Box G

Item	**Excellent**	**tick**	**Good**	**tick**	**Fair**	**tick**	**Poor**	**tick**
Was the percentage of missing items given?	Percentage of missing items described		Percentage of missing items NOT described		-		-	
Was there a description of how missing items were handled?	Described how missing items were handled		Not described but it can be deduced how missing items were handled		Not clear how missing items were handled		-	
Was the sample size included in the analysis adequate?	7* #items and ≥100		5* #items and ≥100 OR 5-7* #items but <100		5* #items but <100		<5* #items	
Were both the original language in which the instrument was developed, and the language in which the instrument was translated described?	Both source language and target language described		-		-		Source language NOT known	
Was the expertise of the people involved in the translation process adequately described? e.g. expertise in the disease(s) involved, expertise in the construct to be	Expertise of the translators described with respect to disease, construct, and language		Expertise of the translators with respect to disease or construct poor or not described		Expertise of the translators with respect to language not described		-	

measured, expertise in both languages								
Did the translators work independently from each other?	Translators worked independent		Assumable that the translators worked independent		Unclear whether translators worked independent		Translators worked NOT independent	
Were items translated forward and backward?	Multiple forward and multiple backward translations		Multiple forward translations but one backward translation		One forward and one backward translation		Only a forward translation	
Was there an adequate description of how differences between the original and translated versions were resolved?	Adequate description of how differences between translators were resolved		Poorly or NOT described how differences between translators were resolved		-		-	
Was the translation reviewed by a committee (e.g. original developers)?	Translation reviewed by a committee (involving other people than the translators, e.g. the original developers)		Translation NOT reviewed by (such) a committee		-		-	
Was the instrument pre-tested (e.g. cognitive interviews) to check interpretation, cultural relevance of the translation, and ease of comprehension?	Translated instrument pre-tested in the target population		Translated instrument pre-tested, but unclear if this was done in the target population		Translated instrument pre-tested, but NOT in the target population		Translated instrument NOT pre-tested	
Was the sample used in the pre-test adequately described?	Sample used in the pre-test adequately described		-		Sample used in the pre-test NOT (adequately) described		-	
Were the samples similar for all characteristics except language and/or cultural background?	Shown that samples were similar for all characteristics except language /culture		Stated (but not shown) that samples were similar for all characteristics except language /culture		Unclear whether samples were similar for all characteristics except language /culture		Samples were NOT similar for all characteristics except language /culture	
Were there any important flaws in the	No other important methodological flaws in the		-		Other minor methodological flaws in		Other important methodological flaws in the design	

design or methods of the study?	design or execution of the study				the design or execution of the study		or execution of the study	
Was confirmatory factor analysis performed?	Multiple-group confirmatory factor analysis performed		-		-		Multiple-group confirmatory factor analysis NOT performed	

Criterion validity: Box H

Item	Excellent	tick	Good	tick	Fair	tick	Poor	tick
Was the percentage of missing items given?	Percentage of missing items described		Percentage of missing items NOT described		-		-	
Was there a description of how missing items were handled?	Described how missing items were handled		Not described but it can be deduced how missing items were handled		Not clear how missing items were handled		-	
Was the sample size included in the analysis adequate?	Adequate sample size (≥100)		Good sample size (50-99)		Moderate sample size (30-49)		Small sample size (<30)	
Can the criterion used or employed be considered as a reasonable 'gold standard'?	Criterion used can be considered an adequate 'gold standard' (evidence provided)		No evidence provided, but assumable that the criterion used can be considered an adequate 'gold standard'		Unclear whether the criterion used can be considered an adequate 'gold standard'		Criterion used can NOT be considered an adequate 'gold standard'	
Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		-		Other minor methodological flaws in the design or execution of the study		Other important methodological flaws in the design or execution of the study	
for continuous scores: Were correlations, or the area under the receiver operating curve calculated?	Correlations or AUC calculated		-		-		Correlations or AUC NOT calculated	
for dichotomous scores: Were sensitivity and specificity determined?	Sensitivity and specificity calculated		-		-		Sensitivity and specificity NOT calculated	

Responsiveness: Box I

Item	Excellent	tick	Good	tick	Fair	tick	Poor	tick
Was the percentage of missing items given?	Percentage of missing items described		Percentage of missing items NOT described		-		-	
Was there a description of how missing items were handled?	Described how missing items were handled		Not described but it can be deduced how missing items were handled		Not clear how missing items were handled		-	
Was the sample size included in the analysis adequate?	Adequate sample size (≥100)		Good sample size (50-99)		Moderate sample size (30-49)		Small sample size (<30)	
Was a longitudinal design with at least two measurements used?	Longitudinal design used		-		-		No longitudinal design used	
Was the time interval stated?	Time interval adequately described		-		-		Time interval NOT described	
If anything occurred in the interim period (e.g. intervention, other relevant events), was it adequately described?	Anything that occurred during the interim period (e.g. treatment) adequately described		Assumable what occurred during the interim period		Unclear or NOT described what occurred during the interim period		-	
Was a proportion of the patients changed (i.e. improvement or deterioration)?	Part of the patients were changed (evidence provided)		NO evidence provided, but assumable that part of the patients were changed		Unclear if part of the patients were changed		Patients were NOT changed	
For constructs for which a gold standard was not available:								
Were hypotheses about changes in scores formulated a priori (i.e. before data collection)?	Hypotheses formulated a priori		-		Hypotheses vague or not formulated but possible to deduce what was expected		Unclear what was expected	
Was the expected direction of correlations or mean differences of the change scores of instruments included in these hypotheses?	Expected direction of the correlations or differences stated		Expected direction of the correlations or differences NOT stated		-		-	

Were the expected absolute or relative magnitudes of correlations or mean differences of the change scores of instruments included in these hypotheses?	Expected magnitude of the correlations or differences stated		Expected magnitude of the correlations or differences NOT stated		-		-	
Was an adequate description provided of the comparator instrument(s)?	Adequate description of the constructs measured by the comparator instrument(s)		-		Poor description of the constructs measured by the comparator instrument(s)		NO description of the constructs measured by the comparator instrument(s)	
Were the measurement properties of the comparator instrument(s) adequately described?	Adequate measurement properties of the comparator instrument(s) in a population similar to the study population		Adequate measurement properties of the comparator instrument(s) but not sure if these apply to the study population		Some information on measurement properties (or a reference to a study on measurement properties) of the comparator instrument(s) in any study population		NO information on the measurement properties of the comparator instrument(s)	
Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		-		Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another construct)		Other important methodological flaws in the design or execution of the study	
Were design and statistical methods adequate for the hypotheses to be tested?	Statistical methods applied appropriate		-		Statistical methods applied NOT optimal		Statistical methods applied NOT appropriate	
For constructs for which a gold standard was available:								

Were hypotheses about changes in scores formulated a priori (i.e. before data collection)?	Hypotheses formulated a priori		-		Hypotheses vague or not formulated but possible to deduce what was expected		Unclear what was expected	
Was the expected direction of correlations or mean differences of the change scores of instruments included in these hypotheses?	Expected direction of the correlations or differences stated		Expected direction of the correlations or differences NOT stated		-		-	
Were the expected absolute or relative magnitude of correlations or mean differences of the change scores of instruments included in these hypotheses?	Expected magnitude of the correlations or differences stated		Expected magnitude of the correlations or differences NOT stated		-		-	
Was an adequate description provided of the comparator instrument(s)?	Adequate description of the constructs measured by the comparator instrument(s)		-		Poor description of the constructs measured by the comparator instrument(s)		NO description of the constructs measured by the comparator instrument(s)	
Were the measurement properties of the comparator instrument(s) adequately described?	Adequate measurement properties of the comparator instrument(s) in a population similar to the study population		Adequate measurement properties of the comparator instrument(s) but not sure if these apply to the study population		Some information on measurement properties (or a reference to a study on measurement properties) of the comparator instrument(s) in any study population		NO information on the measurement properties of the comparator instrument(s)	
Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		-		Other minor methodological flaws in the design or execution of the study (e.g. only		Other important methodological flaws in the design	

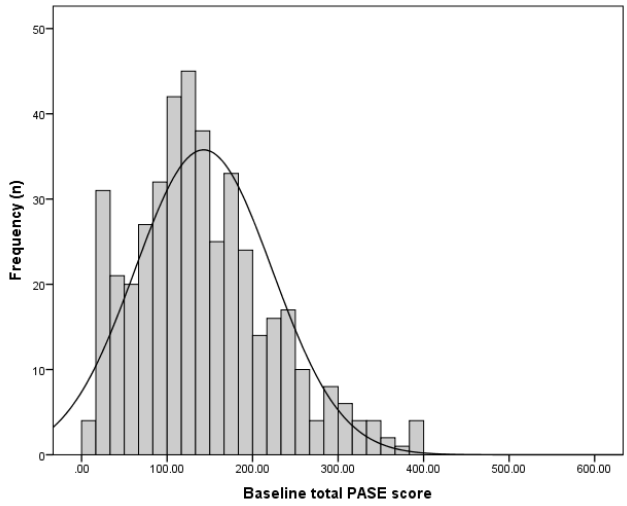
					data presented on a comparison with an instrument that measures another construct)		or execution of the study	
Statistical methods								
Were design and statistical methods adequate for the hypotheses to be tested?	Statistical methods applied appropriate		-		Statistical methods applied NOT optimal		Statistical methods applied NOT appropriate	
For constructs for which a gold standard was available: Can the criterion for change be considered as a reasonable gold standard?	Criterion used can be considered an adequate 'gold standard' (evidence provided)		No evidence provided, but assumable that the criterion used can be considered an adequate 'gold standard'		Unclear whether the criterion used can be considered an adequate 'gold standard'		Criterion used can NOT be considered an adequate 'gold standard'	
Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		-		Other minor methodological flaws in the design or execution of the study		Other important methodological flaws in the design or execution of the study	
for continuous scores: Were correlations between change scores, or the area under the Receiver Operator Curve (ROC) curve calculated?	Correlations or Area under the ROC Curve (AUC) calculated		-		-		Correlations or AUC NOT calculated	
for dichotomous scales: Were sensitivity and specificity (changed versus not changed) determined?	Sensitivity and specificity calculated						Sensitivity and specificity NOT calculated	

Appendix 1.10 SPSS Syntex of VARCOMP

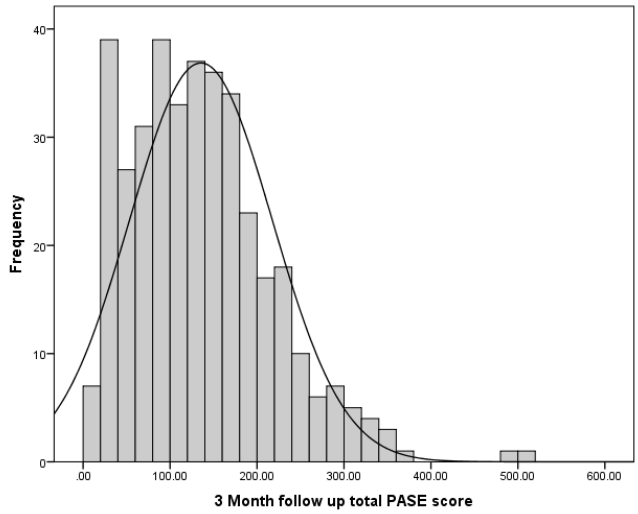
```
COMPUTE dif1=TEST_SCORE-RETEST_SCORE.
COMPUTE mean1=MEAN(TEST_SCORE,RETEST_SCORE).
DESCRIPTIVES VARIABLES=dif1 mean1
/STATISTICS=MEAN STDDEV .
GRAPH /SCATTERPLOT(BIVAR)=mean1 WITH dif1.

* From the Descriptives output use the Mean and StdDev:

COMPUTE Lower=0.0948-(1.96*3.27036).
COMPUTE Upper=0.0948+(1.96*3.27036).

lower = 0.0948 - 6.4099056 = -6.3151056

upper = 0.0948 + 6.4099056 = 6.5047056



VARSTOCASES
 /MAKE rating FROM TEST_SCORE RETEST_SCORE
 /INDEX = test_retest(2)
 /KEEP = SurveyID
 /NULL = DROP.

TEST_SCORE
RETEST_SCORE

VARCOMP
 rating BY SurveyID test_retest
 /RANDOM = SurveyID test_retest
 /METHOD = REML
 /DESIGN = SurveyID test_retest
 /INTERCEPT = INCLUDE .

```

Appendix 1.11 Histogram of baseline and three month total PASE scores and line of distribution

Baseline total PASE scores

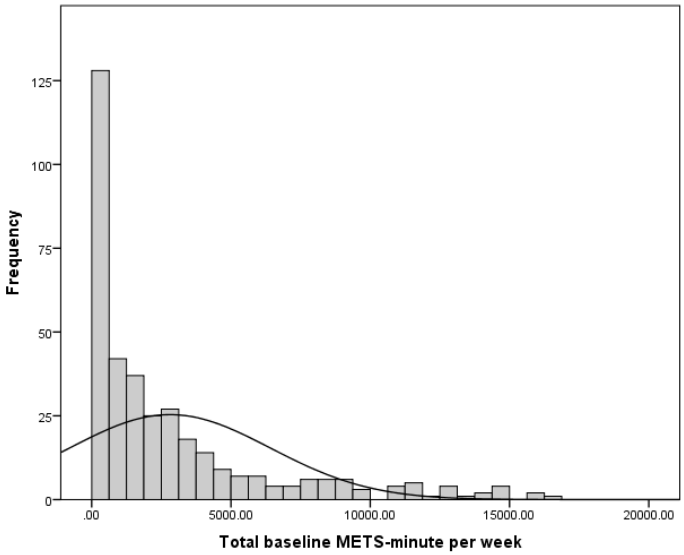


Three month total PASE scores



Appendix 1.12 Histogram of baseline and three month total IPAQ-SF scores and line of distribution

Baseline total IPAQ-SF scores



Three month total IPAQ-SF scores