Benchmarking community and primary care musculoskeletal services: Developing recommendations using evidence syntheses, consensus methods and secondary data analysis

Roanna Mary Burgess

A thesis submitted for the degree of Doctor of Philosophy

June 2022
Keele University

Acknowledgements

I would like to say a huge thank you to my supervision team for their continual support over the last six years. Professor Jonathan Hill: for always being enthusiastic, optimistic, and encouraging, and for putting so much time into providing me with feedback and mentorship. Dr Martyn Lewis: for supporting me so well through my statistical anxieties and always being a calm and considered influence. Dr Annette Bishop: for always keeping me on track and for continuing to provide kind words, opportunities, and advice even after leaving Keele. It has been a pleasure working with you all and I hope to continue to collaborate with you in the future. Thanks also to Dr Gemma Mansell and Dr James Hall for kindly providing support as second reviewers (Chapters 3 & 4 respectively) and for giving advice as friends and colleagues.

I would also like to thank Sandwell and West Birmingham Hospitals NHS Trust (SWBH), with Fiona Shorney, Carolyn Casey, Toby Lewis, and Jim Pollitt helping to make this possible, SWBH Your Trust Charity for financial support, and to the whole of the MSK service for helping me to implement key findings into practice.

Thank you to Versus Arthritis for providing invaluable leadership training through the MSK Champions Programme and to my MSK Champions cohort for peer support and for helping me to celebrate successes and share key outputs.

Lastly, I would like to thank my wonderful family, one of which (who is six now) who was not even born when I started on this PhD journey! Thank you in particular to my husband who has supported me through the ups and the downs and has always encouraged and believed in me, and to my parents for their selfless support of my early education.

Abstract

Introduction: High quality data on service performance is essential in healthcare. There is however a paucity of publicly reported data in community/primary care musculoskeletal (MSK) services. There is also a lack of guidance on which metrics services should be collecting and reporting, and how to adjust data to make fair comparisons across services. This thesis aims to address these gaps, and to develop benchmarking capabilities in this area.

Method: a) a systematic review to identify existing MSK case-mix adjustment models; b) an umbrella review of predictors of MSK functional outcome; c) a systematic review identifying key MSK cost drivers; d) an online survey to develop consensus around a core MSK dataset; e) a secondary analysis of data to test identified case-mix models; f) development of benchmarking recommendations.

Results: Two existing case-mix models were identified from the UK and US. Predictors of MSK function were identified; baseline function, baseline pain severity, mental wellbeing, comorbidities, age and body mass index. Key community/primary care cost drivers were identified; visits to GP, Physiotherapy, and Medical Specialists. Consensus on a MSK core dataset was captured from 166 healthcare professionals and 25 patients across the UK. Secondary analysis of a primary care cohort testing modified versions of the two existing case-mix models showed the US model gave slightly higher predictive power than the UK model (44% and 41% respectively). Finally, the thesis findings were triangulated to develop data collection recommendations for future MSK service benchmarking.

Conclusions: This thesis has generated a body of evidence to inform community/primary care MSK service benchmarking and provides recommendations for future routine data collection of MSK metrics, including; complexity factors for case-mix adjustment, demographics, clinical factors, PROMs, PREMs, and optional cost indicators. The next steps

involve the provision of support and guidance for services to successfully implement these recommendations into practice.

List of Abbreviations

ARMA	The Arthritis and Musculoskeletal Alliance
ARUK	Arthritis Research UK
BMA	British Medical Association
BMI	Body Mass Index
BSR	British Society of Rheumatology
CASP	Critical Appraisal Skills Programme
COMET	Core Outcome Sets in Effectiveness Trials
COS	Core Outcome Set
DoH	Department of Health
DRG	Diagnosis Related Group
EPR	Electronic Patient Record
FABQ	Fear Avoidance Beliefs Questionnaire
FCP	First Contact Practitioner
FE	Fixed Effects
FOTO	Focus on Therapeutic Outcomes Inc.
FS	Functional Status
FYFV	Five Year Forward View
GH	General Health
GLS	Generalised Least Squares
HEE	Health Education England
HES	Hospital Episode Statistics
HRG	Health Related Group
IMD	Index of Multiple Deprivation
LBP	Low Back Pain
MSK	Musculoskeletal
MSK-HQ	Musculoskeletal Health Questionnaire

NEIAA	National Early Inflammatory Arthritis Audit
NHS	National Health Service
NHSE	NHS England
NIHR	National Institute for Health Research
NJR	National Joint Registry
NPROMS	National PROMS Programme
NPRS	Numeric Pain Rating Scale
OA	Osteoarthritis
OHS	Oxford Hip Score
OKS	Oxford Knee Score
OLS	Ordinary Least Squares
OMERACT	Outcome Measures for Rheumatology
ONS	Office of National Statistics
P4P	Payment for Performance
PbR	Payment by Results
PLICS	Patient-Level Information and Costing System
PREMs	Patient Reported Experience Measures
PROMs	Patient Reported Outcome Measures
QoL	Quality of Life
RCT	Randomised Controlled Trial
RE	Random Effects
RI	Research Institute
SES	Socioeconomic status
VA	Versus Arthritis
VAS	Visual Analogue Scale
WPAI	Work Productivity and Activity Index

Table of Contents

List of Tables Included in the Text

List of Figures Included in the Text

1 Chapter 1: Introduction, Aims, and Objectives.

Introduction

This Chapter sets out background literature supporting the importance of data collection, outcome reporting, and benchmarking within healthcare systems, and how if successful, benchmarking can allow for targeted quality improvement of healthcare services. An overview of methods needed to develop benchmarking processes is then provided. This includes review of the use of Patient Reported Outcome Measures (PROMs), known predictors of functional outcome, case-mix adjustment modelling (adjusting for complexity), efficiency and value, key drivers of healthcare costs, and standardised collection of core musculoskeletal (MSK) metrics. The intended PhD aims, objectives, and proposed methodology are then fully described to provide an overview of how this research project sought to help develop benchmarking capabilities within MSK community and primary care services.

Importance of this Research

The NHS Mandate (2017) lays out the need for NHS transformation, with NHS England supporting leaders to drive forwards real improvements in patient care and patient outcomes. Tackling unwarranted variation is highlighted as a priority objective within both the NHS Mandate (2017), and the Five Year Forward View (FYFV) (NHS England, 2014), aiming to reduce the 'unacceptable' care and quality gap.

Musculoskeletal pain conditions are one of the largest disease groups (alongside mental health disorders) contributing to years lived with disability internationally (Vos et al, 2017). 18.8 million people in the UK suffered with an MSK disorder in 2017 (Versus Arthritis, 2019) and the cost to the UK NHS is estimated at £5 billion each year (Briggs et al, 2020). Currently there is a mismatch between the increasing societal burden of MSK pain conditions, and the appropriate health policy response (Blyth et al, 2019).

Versus Arthritis (VA) (formally Arthritis Research UK (ARUK)) published a specific MSK Recommended Indicator Set in 2016. This indicator set defines a shared UK MSK vision, and reflects agreed objectives for MSK health systems; ensuring comparisons in care are made over time, providing information on quality, and supporting quality improvement initiatives both locally and nationally (ARUK, 2016). Specific data capture across the MSK health system was envisaged to enable and support quality improvement in areas such as; early diagnosis, delivering of coordinated care, and empowering people with arthritis and MSK conditions to self-manage (ARUK, 2016). A number of national audits are already in place to capture this data in specific areas outlined within the indicator set, such as; the National Early Inflammatory Arthritis Audit (NEIAA) in the area of 'Rheumatoid Arthritis' (BSR 2020), the National Hip Fracture Database (NHFD) for the area of 'Fragility Fractures' (Royal College of Physicians, 2017a), the National PROMs Programme hip and knee data for the area of 'Osteoarthritis' (NHS Digital, 2020), with additional data on prevalence and spend reported and compared by NHS RightCare in MSK focus packs (NHS England, 2019).

Despite all of the above work on national data collection over the last 10 years however, there is still no national audit or dataset to collect data and provide intelligence about the primary and community care management of non-inflammatory, non-surgical MSK conditions. For these common MSK problems such as back pain and osteoarthritis, we do not have information relating to 'Musculoskeletal Health Outcomes' set out within the Versus Arthritis Indicator set. This includes data on health gain (e.g. Musculoskeletal Health Questionnaire (MSK-HQ)) or utility scores (e.g. EQ5D), or other key metrics such as the percentage of patients whose work is affected by MSK conditions, and the satisfaction of patients treated for MSK conditions in these settings.

The aim of this PhD programme of research was to develop methods to enable future benchmarking of MSK community and primary care services using routinely collected PROM data, standardised information about caseload complexity, and related information on healthcare utilisation and costs. At present, data describing the variability of clinical outcomes and costs, particularly among community and primary care MSK services, is lacking. Research is needed to develop methods around the standardisation of data collection including development of a core set of MSK metrics/outcome measures to capture routine information on costs and quality, and to identify and validate a parsimonious case-mix model to adjust data for complexity allowing for fair comparisons of MSK quality data to be made. These benchmarking methods can then be used moving forwards to inform healthcare policy on how to compare core MSK outcomes/metrics and identify variation/outliers, and subsequently to tackle variation in care quality as identified within the NHS Mandate (NHS England, 2017).

This PhD used data from the National Institute for Health Research (NIHR) funded; Treatment for Aches and Pains Study (TAPS) also known as the STarT MSK (Subgrouping for Targeted Treatment in MuSculosKeletal conditions) cluster randomised controlled trial (RCT) (ISRCTN15366334), commissioned to investigate the effectiveness of stratified care for adults consulting with MSK pain (data available March 2020, n=1211). This cohort data was used to test case-mix adjustment models in a MSK community/primary care patient population.

Lay Summary

There are currently a number of methods that are used to allow direct comparison of services and clinical care for patients with musculoskeletal conditions, taking into account clinical outcomes alongside key factors that indicate complexity including; patient age, gender, and severity of condition. There is however no clear consensus on what data should be collected and reported in routine musculoskeletal care often provided in physiotherapy /GP specialist

clinics, and what factors should influence analysis of performance and resource use. This PhD looked to identify and systematically evaluate existing benchmarking methods, reviewing complexity variables and predictors of poor musculoskeletal outcome, identifying key musculoskeletal service costs, and gaining consensus around routine data collection including patient reported data on outcomes and experiences. This has allowed for clear recommendations to be developed for future benchmarking of musculoskeletal services with a focus to care provided within Physiotherapy and GP led musculoskeletal services in community and GP settings.

This study complemented other outcomes-based projects already underway within the Research Institute, and used data from the Treatment, Aches and Pains Study (TAPS) for analysis within the PhD Project.

Benchmarking

Collection and analysis of patient data on treatment outcome and experience across health systems is fundamental to healthcare success (Nelson et al, 2016). Providing comparative data on care processes and outcomes allows for systematic comparison and analysis across multiple sites, highlighting variation, and allowing for identification of specific targets for improving the quality of patient care (Nelson et al, 2016). Performance data can then drive providers to improve both patient outcomes and service efficiency (Porter, 2009). There are a number of recent examples in UK healthcare of data registries and national clinical audit programmes that have had a positive impact on quality outcome reporting. These include; the National Joint Registry (NJR) and National PROMs Programme within orthopaedics (Smith et al, 2012, NHS Digital, 2019), the Sentinel Stroke National Audit Programme for stroke care (Royal College of Physicians, 2017b), and the National Bowel Cancer Audit for cancer care (Cornish, 2011). These registries allow for the continual collection, analysis and reporting of evaluation and performance data, in order to focus cyclical quality improvement.

Benchmarking in clinical practice represents a process by which individual providers can compare and share best practice, facilitating continuous quality improvement (Siemens et al, 2017). It is a dynamic process delineating best practice and proposing optimal future performance (Siemens et al, 2017). Benchmarking involves selecting appropriate health status measures, and defining an acceptable benchmark. Benchmarks can be derived from within a dataset (e.g. using the statistical mean of the data) or externally using expert judgement or research (DoH, 2012). Research around benchmarking methodology within healthcare is still limited. There is however an increasing focus on reducing unwarranted variation within healthcare systems (Wind et al, 2017, NHS Mandate, 2017), highlighting the need for clinical benchmarking to provide the robust evidence needed to systematically identify variation and guide healthcare improvement/transformation decision-making.

Wind et al (2017) carried out a scoping study to provide an overview of benchmarking research, and to describe benchmarking study characteristics, methods, frameworks and outcomes. Twenty four studies met eligibility criteria for inclusion into the review. Articles were categorised into those using; pathway benchmarking, institutional benchmarking, benchmark evaluation, and benchmarking using patient registries. The authors concluded that benchmarking processes are limited and still under development within healthcare, and are currently most developed in the areas of oncology and ophthalmology (Wind et al, 2017). One successful benchmarking project detailing quality improvement was a study reviewing breast cancer care by Brucker et al, (2008), which demonstrated clear improvement in outcomes using defined benchmarking processes. Success factors for benchmarking identified within the review by Wind et al (2017), included; developing good and simple indicators, voluntary and anonymous participation, involvement of stakeholders, measurement of both quantitative and qualitative data, feedback to clinicians, organised forums for discussion of results, and interpretation by organisational learning not a culture of blame.

Benchmarking Methodology

Patient Reported Outcome Measures (PROMs)

Patient Reported Outcome Measures (PROMs) consist of a series of questions that patients are asked in order to gauge their views on their own health, forming a self-assessment of a patient's health and health related quality of life (Devlin et al, 2010). They provide a standardised method for measuring patient's views (Ahmed, 2012) and therefore allow for collation and comparison. Clinicians use PROM data to guide clinical decision making but PROM data can also be used to evaluate comparative effectiveness when aggregated across patients (Van der Wees et al, 2014). Momentum among policy makers is growing for the routine and mandated collection and reporting of PROMs by clinical services. This standardised information can be used by commissioners and service leaders to aid decision making in relation to resource allocation and highlight best practice and variations in performance (Darzi, 2008). Use of PROMs therefore has the potential to not only empower patients by helping them to make informed decisions, but also to drive forwards quality improvement of healthcare services by allowing for collective analysis (Kyte et al, 2015).

Interpretation of outcome data however needs to give appropriate consideration to extraneous variables that can have a significant impact on overall findings. Known independent variables that can affect outcome include; age, duration of symptoms, surgical history, and medical comorbidities (Deutscher et al, 2009; Rodeghero et al 2015; Werneke et al, 2016). Comparing unadjusted average scores between providers can be misleading as the patient profiles between providers may differ significantly (NHS England, 2012). Adjusting for complexity of case-mix is therefore necessary for making accurate comparisons of provider level care.

Case-Mix Adjustment

Case-mix adjustment is a way of statistically compensating for inter-provider differences in the prevalence of factors that adversely affect treatment response, in order to make between provider comparisons more equitable (Phillips et al, 2003). Historically researchers have utilised regression adjustment to account for differences in measured baseline characteristics between subjects (Austin, 2011). There has also been increased interest in methods based on propensity scores to reduce the effects of confounding/treatment allocation bias in observational data (Austin, 2011). The use of case-mix adjustment methods assumes that researchers can identify the most important prognostic factors of outcome, and that these are appropriately measured and adjusted for. It cannot however address the problems of unknown or un-measurable factors, and therefore it cannot fully eliminate risk of bias (Deeks, et al, 2003). Most case-mix adjustment models are developed using known or theoretically likely independent predictors of outcome, taking into consideration collinearity between variables (Coles, 2010).

Prognostic models are developed to provide estimates of outcome probabilities to support clinical decision making alongside clinical intuition and guidelines (Moons et al, 2009). A good example of this within MSK practice is the Keele STarT Back screening tool which is a prognostic model used to predict patients at risk of a poor outcome and identify sub-groups of patients that may need different approaches to management (Hill et al, 2011). Case-mix adjustment models can involve similar predictive factors, but instead of being used to predict individual outcomes and guide clinical decision making at a patient level, they are used to adjust performance data to allow for service/provider level comparisons/benchmarking. Case-mix adjustment models therefore aim to avoid inclusion of provider variables (e.g. Hospital/clinic type, staffing levels) as these variables could remove effects that may be attributable to local quality improvement initiatives, and potentially can adjust out the differences in quality and performance that are being investigated (Coles, 2010).

Predictors of musculoskeletal treatment outcome

There are a multitude of studies examining prediction of health outcome and prognostic models within MSK health research (Whittle et al, 2017). Many of these cohort studies predict outcome in specific MSK conditions such as low back pain (LBP), shoulder pain, and hip pain, with well conducted systematic reviews summarising agreement on predictive factors (Verkerk et al, 2012, Kooijman et al, 2015, Struyf et al, 2016, de Rooij et al, 2016). Predictive factors emerging from these reviews include; symptom duration, disability level, previous episode, pain severity, baseline PROM score, and level of comorbidity (Kooijman et al, 2015, Struyf et al, 2016, Rooij et al, 2016). Mallen et al (2007) proposed that many of these risk factors were common across pain sites and conducted a systematic review of prognostic factors for generic MSK pain. They found that although there was a high level of heterogeneity across studies, certain generic patient factors were still consistently found to be predictive of outcome. These included pain characteristics (intensity, duration, previous episodes, and multiple pain sites), level of disability, psychological factors such as anxiety or depression, and higher levels of psychological distress (Mallen et al, 2007). Following on from this study, Mallen et al, (2013), carried out a prospective observational cohort study to investigate the predictive ability of generic factors, and found that three preferential generic prognostic indicators emerged; duration of symptoms, pain interference with daily activities, and presence of multiple site pain. These studies support the notion that there are key generic factors across pain sites/MSK conditions that predict treatment outcome.

Previous research on developing MSK case-mix adjustment models

Literature on predictors of outcome in MSK care demonstrates a number of key factors that are consistently found to predict outcome (e.g. baseline severity/PROM score, duration of symptoms, (Kooijman et al, 2015, Struyf et al, 2016, Mallen et al, 2013)) and a number of areas where there are conflicting and inconsistent findings (e.g. age, sex, comorbidities

(Kooijman et al 2015)). Case-mix adjustment modelling involves inclusion of all known patient factors that affect outcome within multivariate modelling, in order for known confounders to be accounted for. Many models include widely accepted confounders, and then add those thought to impact on outcome or with less substantial supporting evidence (Coles, 2010). Variables are then tested independently for effect on outcome, and if significant they are added into multivariable models using a stepwise approach, ensuring as they are added that any additional variables significantly improve model power (Coles, 2010). A number of examples of MSK case-mix adjustment models used to adjust patient outcomes exist, but there is presently no known generic MSK model being used across body regions/conditions and across healthcare settings.

Value and Efficiency in Healthcare

Each year the NHS in England spends in excess of 4.7 billion pounds on MSK healthcare (VA, 2016, Briggs, 2020). Health services globally are also under pressure to deliver more health care with diminishing resources (ARUK, 2016). A central focus on current healthcare delivery is therefore to deliver increasing value for patients (Porter, 2009). Health value can be defined as the health outcome achieved per monetary unit spent (Porter, 2006), with the aim being to efficiently achieve good outcomes, maximising value over the entire care cycle rather than minimising costs by limiting services (Porter, 2009). Widespread PROM collection across the NHS presents an opportunity to develop value-based care, with transformation of services driven by health outcomes achieved per NHS pound spent (Black, 2013, Porter, 2009).

There are a number of studies investigating treatment outcome with a view to efficiency in MSK healthcare. Childs et al (2014) investigated the implications of practice setting on clinical outcomes and efficiency of care in physical therapy services. Efficiency of care was measured simplistically by looking at the difference in number of visits and change in functional status

outcome score achieved between the two settings (hospital outpatient and private practice), giving an overall efficiency score. Deutscher et al (2014) looked at efficiency in relation to clinician post-graduate McKenzie training in a similar way, showing a reduction in clinic visits with those with the highest level of training, supporting the cost-effectiveness of advanced level education. Hart and Connolly (2006) evaluated the feasibility of implementing a 'pay-for-performance' (P4P) or 'value-based purchasing' process in outpatient physiotherapy and occupational therapy. Functional status (FS) change, number of visits, and billing data were reviewed with a view to payment based on effectiveness (FS change) and efficiency (effectiveness combined with number of visits). Reimbursement savings results were illustrated, and results supported the potential positive financial implications of the P4P process.

To increase value within healthcare systems therefore the focus needs to not only be around measuring treatment outcome, but also the accurate measurement of treatment costs and efficiency. Measuring efficiency can involve using simplistic measures such as treatment number or involve more complex approaches designed to cost a full episode of care.

Costing Approaches

There is limited research available evaluating costing of MSK services and evaluating NHS costing methods and resources. Available literature is largely either in the form of economic evaluations within MSK research studies such as RCTs, or in the form of more generic NHS costing guidance and reports. There are a number of different approaches to costing healthcare services, such as a 'top-down' or 'bottom-up' approach, and resources available providing unit costs of health and social care (Curtis, 2016, DoH, 2011). Costing variables taking a health perspective can include; average salary, salary on-costs, qualification and training costs, management, administration and estates costs, capital overheads and travel costs (Curtis, 2016). Costing methodology should depend largely on the intended use of the

information and includes defining the decision problem and objectives of costing, and determining the costing perspective and time horizon (Mogyorosy and Smith, 2005). To develop a costing tool for routine use in clinical practice, the priority needs to be around balancing simplicity and practicality, with accuracy and breadth of information.

MSK Economic Analyses

Economic evaluation is 'the comparative analysis of alternative courses of action in terms of both their costs and consequences' (Drummond et al, 2015). The objectives of health economic analyses are therefore to identify, measure, value, and compare costs and consequences of alternative treatments and programmes of care (Drummond et al, 2015).

There are a number of examples of costing approaches and economic evaluation alongside high quality MSK RCTs. An example includes Hill et al.'s study in (2011), evaluating the use of stratified care models in primary care. This economic evaluation included collecting detail of clinic visits and professionals attended, and assigning value based on best evidence including; Curtis et al (2009); 'Assigning Unit Costs in Health and Social Care', and the NHS Executive (DoH, 2010); 'National Schedule of Reference Costs'. Out of pocket treatments, prescribed medication and work-related absence were also analysed to formulate total healthcare cost scenarios (Hill et al, 2011).

These costing methods are specific, detailed and highly appropriate for collection alongside an RCT, but may be considered too time consuming and unfeasible for collection of cost data in day to day clinical practice. It is therefore important firstly to review the evidence base for costing MSK services, but also to consider the feasibility of data collection by prioritising cost data deemed necessary and practicable for widespread collection across individual MSK services.

Standardising Data

Core outcome sets (COS) are an agreed set of outcomes that should be measured and reported as part of a minimum dataset (Kirkham et al, 2017). They aim to improve comparability across similar trials/datasets, to reduce selective reporting and bias and improve the relevance of data collected in trials and observational studies (Kirkham et al, 2017). The Core Outcome Sets in Effectiveness Trials (COMET) database allows for searching and identification of appropriate COS tools to use in clinical trials and in routine data collections in clinical practice (Williamson et al, 2017).

Examples from the database for MSK include the OMERACT (Outcome Measures for Rheumatology) group who developed through consensus COS measures for use in Rheumatology trials (Williamson et al, 2017), and studies from the ICHOM group that have developed international core outcome sets for low back pain (LBP) and hip and knee osteoarthritis (OA) (Clement et al, 2015, Rolfson et al, 2016).

Formulating recommendations around standardising data for routine collection in MSK clinical practice within community and primary care MSK services forms another important aspect of this PhD programme of research, with consensus sought from MSK stakeholders around what can be feasibly collected in routine care to develop quality data reporting and benchmarking capabilities.

Summary

In summary there is a paucity of research available around benchmarking within healthcare, and a clear mandate within the NHS currently (NHS Mandate, 2017) for a focus on productivity and efficiency gains, tackling unwarranted variation, and ensuring NHS funds are spent on better care and treatments for patients that are sustainable for years to come. Benchmarking within healthcare provides an opportunity to provide the necessary information to transform services based upon both effectiveness of care, and associated costs and efficiency, in order to highlight exemplar service models, and those that require

improvement. This PhD programme of research focused on developing specific methodology for undertaking performance benchmarking within an MSK community/primary healthcare context, to allow for future targeted quality improvement and transformation of MSK services. The aim was to address the key areas of benchmarking where evidence is currently lacking within the field of MSK community/primary care. These areas include case-mix adjustment of MSK PROM scores, costing MSK services, and identifying a core MSK dataset for capture across routine MSK care. Research findings were then used to formulate recommendations alongside discussion of barriers and enablers for embedding this in routine healthcare. Review of the literature highlighted a particular scarcity of research within a UK community and primary care setting. Data for analysis within this programme of research therefore focused on UK MSK community and primary care services with a view to developing a methodology for this area that could then be tested further across an MSK pathway of care.

Aims and Objectives

Aims: The overall aim of the PhD programme of research was to provide guidance on which metrics community/primary care MSK services should be collecting and reporting, on how to adjust this data in order to make fair comparisons between services, and on how these findings could underpin a national evaluation/audit in this setting, looking to drive improvements in quality and value of MSK care.

Protocol for programme of research (developed a priori)

Specific objectives to be addressed by the PhD proposal:

OBJECTIVE 1: To identify existing models used for case-mix adjustment of PROMs data in MSK services, and to identify any additional candidate variables (not used within existing models) that have strong evidence supporting their use in predicting functional outcome.

Research Questions:

I. What existing models are available to case-mix adjust patient reported outcome measure (PROM) data in a musculoskeletal (MSK) population?
II. Are there any further patient factors that are strong predictors of MSK functional outcome (PROM scores) that could be used for case-mix adjustment?
III. What recommendations can be made for future case-mix adjustment of MSK PROM data?

Method: Systematic review of the literature identifying existing MSK case-mix adjustment models (used to adjust PROM data) and further umbrella review identifying generic predictors of MSK functional outcome (measured using PROMs) to be considered as further candidate variables for use in case-mix adjustment modelling.

A systematic review of the literature will be carried out following PRISMA guidelines, to identify studies that have used case-mix adjustment models to adjust MSK clinical outcomes data. Electronic databases will be searched including AMED, CINAHL, MEDLINE, EMBASE and HMIC, alongside searching the grey literature. Quality of included studies will be assessed using a specific quality tool. Data extraction sheets will be populated to summarise study findings and aid input into the review. A narrative analysis is envisaged to explore key themes within identified models utilised for case-mix adjustment of MSK outcomes, including development history, model validity and predictive ability.

A further umbrella review will then be conducted to ensure no other useful candidate variables have been omitted from existing models. This review will aim to identify any systematic reviews that analyse the ability of individual patient factor variables to predict functional outcome, measured using PROMs. The review will be restricted to time beyond the latest iterations of existing models in order to identify additional variables that may not have

been considered within existing models. The purpose is to add to the evidence already collated within the systematic review.

Following the two reviews, recommendations will be made about case-mix adjustment methods, including which variables to consider including within a generic MSK case-mix adjustment model to be used across MSK conditions.

OBJECTIVE 2: To identify key drivers of healthcare costs for patients presenting to community/primary care with MSK conditions.

Research Question:

I. What key cost variables/indicators can be used to determine costs of community/primary care MSK services?

Method: Review of grey literature alongside a systematic review of economic analyses to identify key cost drivers of MSK healthcare costs.

A scoping literature review will be conducted to explore key 'resource use' variables used for costing community/primary care MSK services, followed by a systematic review of economic analyses set predominantly in community/primary care MSK services to identify the key drivers of MSK healthcare costs. Electronic databases will include Medline, AMED, EMBASE, CINAHL, HMIC, BNI, and HBE. Resource utilisation, resource unit costs and mean resource use costs will be extracted and reported for the usual care/control groups. This review will be descriptive in nature with the aim of identifying the key resource use variables driving MSK costs, helping to formulate a list of potential cost variables to be considered and prioritised within a stakeholder consensus process for inclusion within a standardised MSK dataset.

OBJECTIVE 3: To gain consensus from key stakeholders over what outcome measures and metrics should be included within a standardised MSK dataset to capture/report routine quality data within community/primary care MSK services.

Research Question:

I. What metrics need to form part of a minimum dataset for routine reporting of MSK data?

II. What are the current challenges to collecting an MSK core outcome set (COS)?

Method: Online consensus survey involving multiple stakeholders (including both clinicians and service users) providing feedback on a core outcome set/routine collection of MSK metrics.

A consensus process will be developed to determine the minimum dataset requirement for effective and useful capture and analysis of MSK data. This will be achieved through an online consensus survey.

Potential MSK metrics identified following completion of objectives 1 and 2 will be listed with detail of why they are included for consideration in a minimum dataset/core outcome set. Alongside this, widely used generic PROM tools and PREM domains will be listed to develop consensus around what should be included in a minimum dataset that could be used for MSK benchmarking in this setting. Ethical approval will be sought detailing the content of the online survey. Stakeholders will include both healthcare professionals and service users. The survey will be developed using Lime Health Survey software provided by the University.

OBJECTIVE 4: To explore through secondary analysis of data the predictive ability of existing case-mix adjustment models in predicting PROM scores in a UK community/primary care MSK population and investigate other identified patient factors/variables (identified from objective 1) using univariate and multivariate regression modelling.

Research Questions:

I. How useful are existing MSK case-mix adjustment models in predicting functional outcome (PROM scores (MSK-HQ)) in a UK MSK community/primary care cohort?
II. Are there other patient factor variables that should be included within a generic UK MSK case-mix adjustment model for use in community/primary care?
III. What is the most parsimonious MSK case-mix model that could be feasibly adopted into practice in community and primary care services?

Method: Univariate and multivariate regression analysis of secondary dataset (STarT MSK RCT data (n=1211) (RI funded study))

Using results from the systematic review and from the additional umbrella review, an initial list of evidence-based patient factor variables feasible for widespread clinical collection will be formulated. Available variables will then be explored within a community/primary care patient cohort (STarT MSK Main Trial Data (n=1211, data available March 2020)).

Variables will be analysed to explore independent predictive ability, and collective predictive ability, using univariate and multivariate regression modelling respectively. Existing case-mix adjustment models will be tested (using all available variables therefore likely requiring minor modifications). Comparisons will then be made with regards to predictive ability of existing case-mix adjustment models used within this UK community/primary care cohort, compared to an evidence informed model using all of the recommended available variables identified in objectives 1 and 2, and a statistically informed model developed using all available variables within the STarT MSK trial cohort dataset. Evaluation will include considering model predictive ability against model feasibility to make recommendations for future practice.

The main practical issue with secondary analysis of existing data will be the availability of identified independent variables. It is therefore proposed that on confirmation of agreed

variables from the systematic review and additional umbrella review, opportunities for new data collection to include additional baseline variables will be identified, and these variables included where possible within existing funded studies (STarT MSK Main Trial).

OBJECTIVE 5: To formulate recommendations for the development of MSK benchmarking in community/primary care MSK services, providing the framework for a national MSK evaluation/audit.

Research Question:

I. What recommendations can be made for developing benchmarking capabilities and working towards a national MSK evaluation/audit of community and primary care services?

Method: Overview of findings from previous chapters to develop detailed recommendations to outline what would be needed to take this to a national data collection/benchmarking programme.

Findings and recommendations from each of the thesis chapters will be brought together in a narrative review, detailing evidence-based recommendations for the development of benchmarking methods/national audit capabilities in UK community/primary care MSK services and identifying barriers and enablers to adopting these recommendations into practice.

Plan for Thesis Chapters

Chapter 1: Introduction, Aims and Objectives.

Chapter 2: Models used for case-mix adjustment of PROMs data in musculoskeletal services: A systematic review of the literature.

Chapter 3: Identification of patient factor variables for use in case-mix adjustment of musculoskeletal PROM data.

Chapter 4: Costing community and primary care musculoskeletal services.

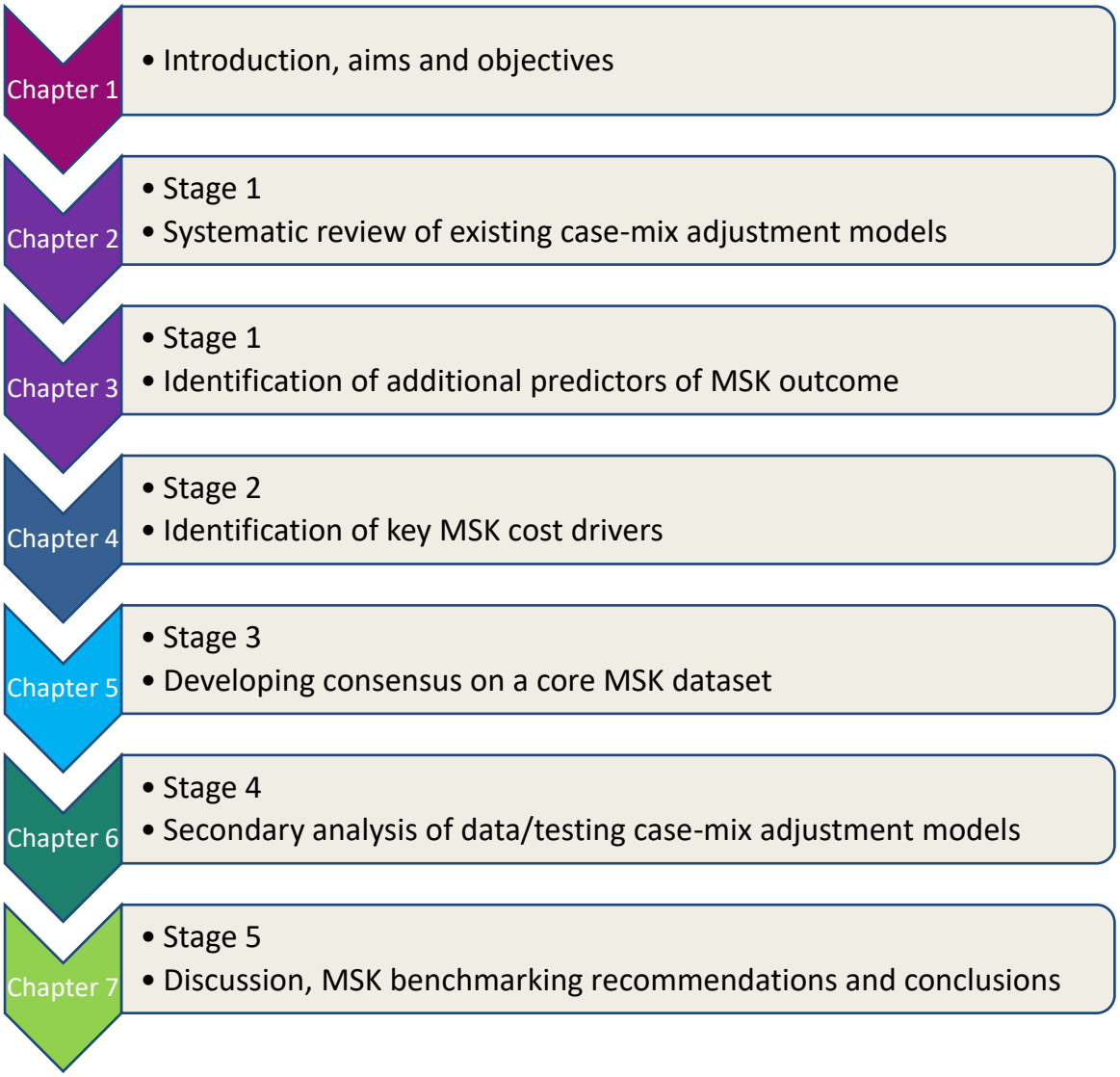
Chapter 5: Developing a core outcome set for community and primary care musculoskeletal services: A consensus approach.

Chapter 6: Secondary analysis of data to investigate the predictive ability of case-mix adjustment models/variables in a community/primary care cohort.

Chapter 7: Recommendations for future MSK benchmarking in community/primary care.

This chapter sets out a detailed plan for achieving five specific PhD objectives, in order to determine the optimum methodology for the future benchmarking of MSK services (see **Figure 1-1** for summary). Each objective has been explored in full within the following core chapters of the thesis (**Chapters 2 to 6**), and within the final summarising chapter (**Chapter 7**), and a detailed outline proposal for developing future national MSK benchmarking supported by findings from each stage of the PhD project has been provided (**Chapter 7**).

Figure 1-1 Plan of Thesis Chapters



Publications and Presentations

I. Burgess, R., Bishop, A., Lewis, M. and Hill, J., 2019. Models used for case-mix adjustment of patient reported outcome measures (PROMs) in musculoskeletal healthcare: A systematic review of the literature. *Physiotherapy*, *105*(2), pp.137-146

Presented at PROMS UK 2018 (platform presentation) and PGR Symposium 2018 (poster presentation)

II. Burgess, R., Mansell, G., Bishop, A., Lewis, M. and Hill, J., 2020. Predictors of functional outcome in musculoskeletal healthcare: An umbrella review. *EJoP*, *24*(1), pp.51-70

Presented at PROMS UK 2019 (poster presentation)

III. Burgess, R., Hall, J., Bishop, A., Lewis, M. and Hill, J., 2020. Costing Methodology and Key Drivers of Health Care Costs Within Economic Analyses in Musculoskeletal Community and Primary Care Services: A Systematic Review of the Literature. Journal of primary care & community health, 11, p.2150132719899763

Presented at Physiotherapy UK 2020 (platform presentation) and PGR Symposium 2019 (3 minute thesis, winning best systematic review)

IV. Burgess, R., Lewis, M., McRobert, C. and Hill, J.C., 2021. Developing a Core Outcome Set for Community and Primary Care Musculoskeletal Services: A Consensus Approach. *Musculoskeletal Science and Practice*, p.102415.

Presented as part of a Physiotherapy UK 2021 Focused Symposium, and as a poster at Physiotherapy UK 2021.

V. Burgess, R., Lewis, M., and Hill, J.C., 2021. Musculoskeletal case-mix adjustment in a UK primary/community care cohort: Testing Musculoskeletal models to make recommendations in this setting. *Musculoskeletal Science and Practice,* https://authors.elsevier.com/sd/article/S2468-7812(21)00139-9

Presented as a poster at Physiotherapy UK 2021.

2 Chapter 2: Models used for case-mix adjustment of PROMs data in musculoskeletal services: A systematic review of the literature.

This Chapter presents a systematic review of the literature to identify existing case-mix adjustment models used to adjust MSK PROM data.

Abstract

Background: Case-mix adjustment is an established method to take account of variations across cohorts in baseline patient factors, when comparing health outcomes. Although commonplace, there is a lack of evidence as to the most appropriate case-mix adjustment model to use to enable fair comparisons of PROM data in musculoskeletal services.

Objectives: To conduct a systematic review summarising evidence of the development, validation, and performance of musculoskeletal case-mix adjustment models, and to make recommendations for future methods.

Data Sources: Searches included; AMED, CINAHL, EMBASE, HMIC, MEDLINE, and grey literature.

Eligibility Criteria: Studies; from January 1992-May 2017, English language, musculoskeletal adult population, developing or validating a case-mix adjustment model, using a relevant PROM, and using patient factors feasible for clinical collection.

Data Synthesis: Two reviewers evaluated selected papers. The CASP Cohort Tool was used to assess quality.

Results: Fourteen studies were included; eight US studies on the Focus on Therapeutic Outcomes model (pooled n=546,726 patients (with pre/post treatment data)) and six UK studies related to the UK National PROMs Programme model (pooled n=282,424 patients (with pre/post treatment data)). The majority used retrospective data, restricted to complete datasets. Both US and UK models showed good predictive ability (R2 18-42%). Common model variables were; baseline PROM score, age, sex, comorbidities, symptom duration, and surgical history. Reduced quality

scores were mainly due to acceptability of patient recruitment, and completeness and length of patient follow up.

Conclusion: Significant methodological crossover was found. Further studies are however needed to externally validate and develop models across MSK settings.

Contribution to the Thesis:

- This systematic review has identified two broad MSK case-mix adjustment models, and highlights both the commonalities in case-mix adjustment approaches but also the need for further good quality studies to inform future practice.
- Effective case-mix adjustment modelling across MSK clinical pathways of care will allow for further development of performance profiling and benchmarking across MSK practice, with the aim of improving quality and equity of MSK healthcare provision.

The contents of this chapter have been published in part in:

Burgess, R., Bishop, A., Lewis, M. and Hill, J., 2019. Models used for case-mix adjustment of patient reported outcome measures (PROMs) in musculoskeletal healthcare: A systematic review of the literature. *Physiotherapy*, *105*(2), pp.137-1

Introduction

Routine use of patient reported outcome measures (PROMs) can help patients and clinicians make better decisions and enable comparisons of providers' performance facilitating quality improvement (Black, 2013). An example of this is the UK National PROMs Programme which has successfully raised standards in the area of hip and knee replacement surgery (NHS England (NHSE), 2016). Patient outcomes are a function of; therapeutic intervention effectiveness, quality of care, patient attributes or 'risk factors' affecting response to care, random variation, and the natural course of a condition (Iezzoni, 2009, Vasseljen et al, 2013). Case-mix or risk-adjustment (termed case-mix adjustment here for consistency) is a statistical process that aims to account for differences in the mix of patient attributes across definitive patient cohorts, in order to make fair comparisons of the relative effectiveness (outcome) of care provided (Iezzoni, 2009). For example, to enable fair comparisons across different MSK physiotherapy services it may be appropriate to adjust for differences in age or symptom duration, as older patients whose symptoms are more chronic are likely to report less functional status change than those that are younger and have more acute symptoms (Hart et al, 2011). Other factors for consideration when adjusting health outcomes may include patient variables such as; gender, symptom severity, and impairment type (Hart and Connolly, 2006). These factors can be described as pre-existing baseline patient factors as they are completely beyond the control of the provider, unlike provider factors such as the clinic setting or treating clinician, which also influence health outcomes (Werneke and Hart, 2001). Case-mix adjustment aims to avoid inclusion of provider variables as these variables could remove effects that may be attributable to local quality improvement initiatives, and potentially can adjust out the differences in quality and performance that are being investigated (Coles, 2010). For example, if one physiotherapy service used only advanced physiotherapy practitioners and another employed only basic grade non specialist staff, and grade of treating clinicians was adjusted for then any difference in quality of these services due to the differing skill-mix would be adjusted out rather than being used to help

explain differences in quality, and guide quality improvement initiatives. Most models therefore only adjust for patient factors to allow for fair inter provider comparison, and also because these variables are feasible for widespread collection in clinical practice as they are most often patient reported or are routinely available in healthcare records (Coles, 2010).

Within an MSK context the evidence for existing case-mix adjustment models comparing provider level health outcome data has not been systematically evaluated, and there has been no previous review of the literature to the authors' knowledge. This review therefore aims to summarise the evidence for the development, validation, and performance of MSK case-mix adjustment models, and make recommendations for future case-mix adjustment methodology.

Methods

This review followed protocol guidance set out within the PRISMA statement (Moher et al, 2015), and has been registered on the PROSPERO database (CRD42017055948).

ELIGIBILITY CRITERIA:

Inclusion criteria for the review were: studies from 1992 to May 2017 in line with early implementers of widespread MSK PROM collection (FOTO Inc., 2018) and to allow for currency and applicability of results, studies reported in the English language due to resource limitations, observational cohort studies, cohorts of adult patients seeking treatment for MSK conditions, use of a case-mix adjustment model (focus on development, refinement or validation), case-mix adjustment of PROM data, PROM data capturing treatment effect/change (not at a single point in time), and models including variables feasible for widespread clinical collection (not using variables such as imaging results that would not be uniformly collected). Exclusions were studies not reporting detailed results, and those that failed to include statistical data reporting model effectiveness (see **Table 2-1**).

Table 2-1 Inclusion and Exclusion Criteria

Inclusion	Justification	Exclusion
Studies from 1992 to current date	Applicability to current case-mix adjustment theory/best practice. The US Focus On Therapeutic Outcomes (FOTO) Incorporation were early implementers of physical therapy outcomes and case-mix adjustment theory. FOTO was founded in 1992, with case-mix adjustment fully implemented within the system by 2003 (FOTO Inc., 2018).	Studies prior to 1992
Observational studies (e.g. cohort, case series, secondary use of RCT data only)	No interventional/comparator studies. Only studies collecting outcome level data alongside baseline variables in order to develop or validate a case-mix adjustment model.	RCTs and other interventional studies
Use of a case-mix adjustment model including development, refinement, or validation of existing models.	Must use a case-mix adjustment model to allow for comparison between models and review of best practice. Model development, refinement or validation must be one of the primary objectives of included studies.	No case-mix adjustment model used in study, or case-mix adjustment used to adjust data but not the focus of the study.
Musculoskeletal (MSK) setting	Collecting relevant variables for optimum MSK case-mix adjustment tool	Non MSK settings
Models must focus on adjustment of PROMs (PROMs used in MSK setting such as HRQoL and FS alongside disease specific PROMs such as Oxford Knee Score and Oxford Hip Score)	Case-mix adjustment variables and prediction models for specific outcome such as mortality would not be transferable to prediction of general day to day MSK clinical PROMS	Not focused on prediction of specific medical outcome or complications such as mortality or adverse events (Devlin et al, 2010)
Study must be looking at case-mix adjustment of treatment/intervention outcome data (health gain/change over time) not just at case-mix adjustment of functional status	Need to be evaluating treatment outcome to inform best practice of a case-mix adjustment model that could be used to benchmark MSK services	Studies using case-mix adjustment with only one PROM capture or not carrying out intervention between PROM capture points

Variables included in model must be feasible to be collected in everyday clinical practice	Review aims to make recommendations for future modelling in day-to-day clinical practice to inform performance profiling therefore variables must be feasible for widespread collection.	Variables included in the model impractical for day to day collection in clinical practice
Studies must include adult population	Studies focussing on paediatric population would not be transferable to adult MSK services	Studies on paediatric populations
Studies must report model effectiveness using either R2 statistic, individual significance of variables within models using beta coefficients, or variable odds ratios.	Study must give information about effectiveness of the case-mix adjustment model to inform best practice	Studies not reporting detailed results to include statistical data to support model effectiveness

FS; functional status, HRQoL; health related quality of life, MSK: musculoskeletal, PROMs; patient reported outcome measures.

SEARCHES:

A search strategy was developed iteratively with guidance from a health librarian with expertise in systematic review searching, initially conducting test searches for a single database until the refined strategy was agreed. This involved amalgamating sets of search terms, reducing individual terms, and exploding terms such as 'musculoskeletal' to optimise the balance between sensitivity and precision of searches undertaken (Higgins and Green, 2011). Search terms included key words for; target population, MSK conditions, outcomes, and methodology (see **Table 2-2**). Full searches included electronic databases of: CINAHL, MEDLINE, EMBASE, AMED and HMIC (see **Appendix 2-1** for full search strategy) from Jan 1992 to May 2017. Grey literature included searches of NHS Evidence websites of the Department of Health (DoH, 2017) and NICE (NICE, 2017). Additional searches included searching reference lists and use of citation tracker for included studies identified from electronic database searches. Seminal authors/research groups were also contacted for all identified models to ensure latest iterations of models were included. Further International seminal published authors within the PROMs literature were contacted for

knowledge of any other case-mix adjustment models or papers for consideration and included authors from Australia, US, Netherlands and UK.

Table 2-2 Table of search terms

Population	Treatment of (MSK Condition)	Outcome	Study Design (methodology)
Physiotherapy	Low back	Patient Reported Outcome Measure (PROM)	Case mix adjustment
Physical Therapy	Cervical	Effectiveness	Risk adjustment
Rheumatology	Spine (Spinal)	Change score	Regression analysis
Orthopaedics	Hip	Health gain	
Chiropractic	Knee	Functional status	
Osteopathic	Shoulder	Quality of Life	
Rehabilitation	Musculoskeletal		

SELECTION PROCESS:

One independent reviewer (RB) undertook a preliminary screen of all titles to remove studies clearly and unquestionably excluded from the study. RB then screened all remaining abstracts identified from searches alongside a second reviewer (AB or JH). Two independent reviewers (RB and JH or AB or ML) then read full articles identified to see if they met all the inclusion criteria. RB read all articles, and JH, AB and ML had articles divided between them.

DATA EXTRACTION AND QUALITY ASSESSMENT:

Information on identified articles was independently entered onto a data extraction form by the two reviewers, with the form reflecting the key themes from the STROBE Checklist (Von Elm et al, 2007), and quality assessed using the CASP Cohort Quality Tool (CASP, 2017). Agreement on study inclusion was discussed between RB and the second reviewers. Any study without agreement between the two reviewers was able to be discussed between all reviewers until agreement could be made. All studies were however agreed between two reviewers without disputes.

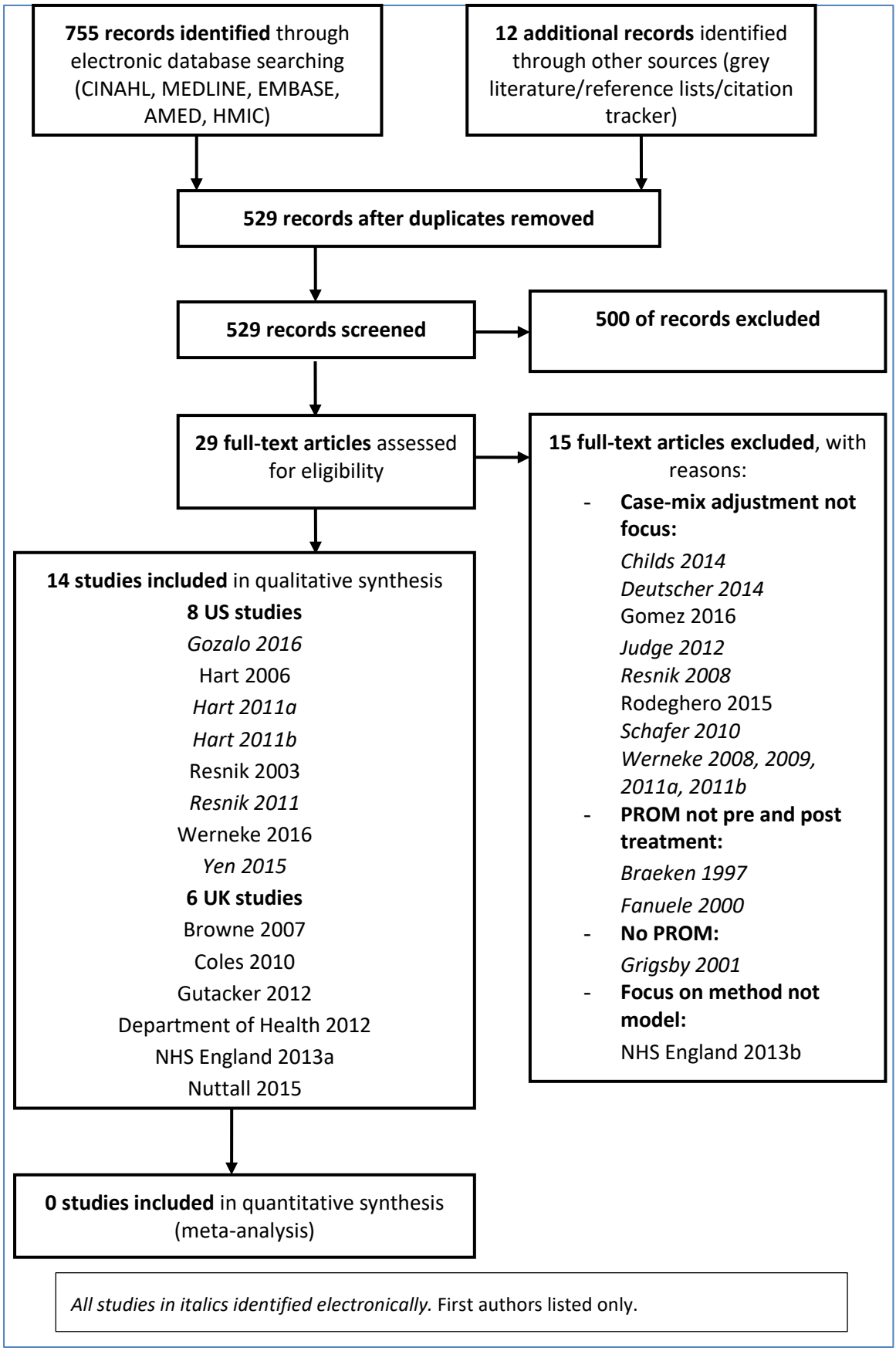
A systematic narrative synthesis was conducted, with information presented in table and text format to summarise and explain the history and development of identified case-mix adjustment models, and the overall study findings. It was not possible to perform a meta-analysis of the studies by pooling data due to the large methodological diversity (heterogeneity) among studies (Higgins and Green, 2011), including using different patient factors across model iterations and different statistical methods. For this reason, results were summarised in tables and discussed in detail. For each case-mix adjustment model associated studies/papers and statistical methods are presented together to allow for ease of viewing overarching study findings.

Results

SEARCH RESULTS:

Electronic database searches identified 755 articles for consideration for inclusion, 517 after duplicates were removed (see **Figure 2.1**). Grey literature and additional searching identified a further 12 articles for consideration. Six of seven experts in the field responded to contact made via email or telephone call and for the expert who did not respond an alternative expert within that research group was found who in turn responded to email contact. This identified one manuscript that was being prepared for submission that is not included within the review. Following screening, fourteen articles were appropriate for inclusion (see **Figure 2-1**). Two broad case-mix adjustment models were identified (US Focus on Therapeutic Outcomes (FOTO), and UK National PROMs (NPROMS)).

Figure 2-1 Flowchart of search results



Eight of the fourteen studies included in the review were undertaken in the US, using data from the FOTO database (Resnik and Hart, 2003; Hart and Connolly, 2006; Hart et al, 2011a; Hart et al, 2011b; Resnik et al, 2011; Yen et al, 2015; Werneke et al, 2016; Gozalo et al, 2016). Four of the eight US papers were authored (primary author) by members of the FOTO Research Advisory Board (FRAB) (Hart and Connolly, 2006; Hart et al, 2011a; Hart et al, 2011b; Werneke et al, 2016). The other four papers were undertaken by independent first authors given access to the FOTO database (Resnik and Hart, 2003, Resnik et al, 2011; Yen et al, 2015; Gozalo et al, 2016). Two of these four studies were also co-authored by members of the FRAB (Resnik and Hart, 2003, Resnik et al, 2011). Participants in included US studies ranged from 323 post exclusion of missing data (Hart et al, 2011b), to 189,088, post exclusions and cleaning of data (Hart et al, 2011a). The pooled sample size across US studies with pre and post treatment data was 546,726.

Six of the fourteen studies meeting inclusion criteria for the review were UK based. These included feasibility work for the NHS England NPROMS Programme (Browne et al, 2007), NPROMS publications (Coles, 2010; DoH, 2012; NHSE, 2013a), and independent researchers using NPROMS data (Gutacker et al, 2012; Nuttall et al, 2015). All of these studies were identified following review of the grey literature/additional searches as they were all NHS publications or secondary analyses of NHS data. Participants in included UK studies ranged from 387 (Browne et al, 2007), to NPROMS data collected from; 2009-10 (85,177), 2010-11 (95,406), 2011-12 (101,454) totalling 282,037 patients (DoH, 2012; NHSE, 2013a, HSCIC, 2014). The pooled sample size across UK studies with pre and post treatment data was 282,424.

Follow up was standardised at six months across UK studies but was non-standardised in US FOTO studies with collection at the end of the treatment episode. All included studies were cohort studies, with three prospectively collecting data (Browne et al, 2007, Coles, 2010; Resnik et al, 2011), and the remaining studies undertaking retrospective analyses of existing datasets.

For results detail see **Table 2-3** for quality of included studies, **Table 2-4** for summary of articles included, and **Table 2-5** for summary of model variables within included studies.

QUALITY APPRAISAL:

Quality of included studies was evaluated using the CASP Cohort criteria. Quality across studies was good using this tool (see **Table 2-3**). There were however consistent sources of bias across US and UK studies within identified areas such as patient recruitment and completeness of follow up, which are discussed below.

Key sources of bias across US studies included: Selection-bias due to the exclusion of those with missing data (see **Table 2-3** for percentage of those with complete data, where available). Hart et al (2011b) for example were only able to include 323 of 39,529 patients (0.8%) within their routine dataset as only these patients had data for all psychosocial measures pertinent to the study, as collecting multiple psychological measures was not routine practice, this however may have significantly biased the sample who were likely to be more psychologically impaired and was acknowledged by the authors. Three of the eight US studies did however use inverse probability weighting to account for missing data (Resnik et al, 2011; Yen et al, 2015; Gozalo et al, 2016), and four studies compared baseline characteristics between those with missing and complete data to assess likelihood of bias, broadly concluding that although some differences were found that these were unlikely to lead to systematic selection biases as missing data included both patients with characteristics associated with better and worse outcome (Hart and Connolly, 2006; Hart et al, 2011a; Hart et al, 2011b; Werneke et al, 2016) limiting effect on predictive models. Patients were however also limited to those attending clinics using FOTO software so may not be representative of clinics across the US (n=4776 clinics currently across the US (FOTO Inc., 2018)). All US studies had non-standardised follow-up outcome assessment time-points with collection at the completion of the individual's treatment episode, both preventing the collection of follow up data for those who ceased attending for treatment and limiting the ability to quantify estimates

of efficacy for a given time. Patients with missing follow up data may therefore be 'missing not at random' (Gomes et al, 2016) as they have chosen to cease attending leading to further potential attrition bias (Higgins and Green, 2011). Resnik and Hart (2003) reported that these patients were younger and had higher functional status scores and therefore supposed that they were likely not returning due to resolution of their symptoms. Not including those with greater chances of improvement will however impact on model predictive ability and therefore overall study findings, as will variation in timing and mode of data collection (Whittle et al, 2017).

Key sources of bias across UK studies included: Selection-bias due to the exclusion of those with missing data (see **Table 2-3**). The study by Browne et al (2007) did use the SF-36 rule (Ware et al, 1993) for dealing with missing data, but 25% of eligible patients were excluded due to failure to invite patients to participate. Due to data linkage between data sources within the NPROMS Programme, unlinked data were also not able to be included in the full analysis, which could again potentially bias the final patient sample. In 2011/12 116,734 of 247,699 patients who underwent PROMS eligible procedures had complete and linked data (47.13%), this was 63.1% of those who completed baseline PROM data (HSCIC, 2014). Whether this impacted on results would depend on whether unlinked or missing data was missing at random (Gomes et al, 2016) or whether this was due to systematic poor administrative processes at certain provider NHS trusts, which is unclear. Follow-up data collection across UK studies was standardised at a six-month time-point reducing likelihood of bias. Collection at treatment commencement however was both through pre-admission clinics and at admission for the PROM procedure, leading to a small source of variation.

All included studies used data from clinical databases and therefore all studies were impacted by the ability to control the quality of the data included and attrition rates of patients providing data for their episode of care. All studies reported these limitations reinforcing the issues around using

clinical data for research purposes, however although acknowledged this will have led to a high risk of bias for this domain within included studies (Higgins and Green, 2011).

Table 2-3 Quality assessment using CASP tool

	CASP Cohort Tool												
Author	**Clearly focussed**	**Recruit-ment acceptable**	**Exposure accurately measured**	**Outcome accurately measured**	**Identified con-founding**	**Accounted for con-founding**	**Subject FU complete enough**	**Subject FU long enough**	**Results Precise**	**Believe results**	**Applicable results**	**Fit with other evidence**	**Complete data**
First Author US													
Resnik 2003	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	
Hart 2006	Yes	CT	Yes	Yes	Yes.	Yes	Yes	Yes	Yes	CT	No	CT	62%
Hart 2011	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	62%
Hart 2011	Yes	No	Yes	Yes	Yes	Yes	No	CT	No	Yes	Yes	CT	0.80%
Resnik 2011	Yes	No	Yes	Yes	Yes	Yes	No	CT	Yes	Yes	Yes	Yes	44.30%
Yen 2015	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Gozalo 2016	Yes	Yes	Yes	Yes	Yes	Yes	No	CT	Yes	Yes	Yes	Yes	57.20%
Werneke 2016	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes	35%
First Author UK													
Browne 2007	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	CT	90.2-91.6%
Coles 2010	Yes	CT	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
DoH 2012	Yes	CT	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
NHSE 2013a	Yes	CT	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Gutacker 2012	Yes	CT	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	
Nuttall 2015	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	83.90%
(FU; follow up, CT; Can't tell)													

Table 2-4 Summary of included studies

Author	Design	Setting	Data Sources	Study Size (complete/included datasets)	PROMs	Number of variables	Model R2 (where available)
US Studies							
Resnik and Hart (2003)	Retrospective cohort	Outpatient physical therapy	FOTO	24,276	OHS, SF-12, SF-36	8	35-42%
Hart and Connolly (2006)	Retrospective cohort	Outpatient therapy	FOTO	189,088	FS	12	35-36%
Hart (2011a)	Retrospective cohort	Outpatient therapy	FOTO	49,376	FS	8	30%
Hart (2011b)	Prospective cohort	Outpatient therapy	FOTO	257	FS	10 (plus PM)	31% (intake model)
Resnik (2011)	Prospective cohort	Outpatient therapy	FOTO	44,925	FS	8	18-40%
Yen (2015)	Retrospective cohort	Outpatient therapy	FOTO	147,623	FS	7	31% (FE model)
Werneke (2016)	Retrospective cohort	Outpatient physical therapy	FOTO	723	FS	13 (tested in BM)	35% (BM)
Gozalo (2016)	Retrospective cohort	Outpatient therapy	FOTO	90,392	FS	8	
UK Studies							
Browne and Black (2007)	Prospective cohort	Orthopaedic		351 hip and 349 knee patients	EQ5D Index, OHS, OKS, SF-36	8	24-27%
Coles (2010)	Prospective cohort	Orthopaedic	NPROMS	14,041 hip and 15,718 knee patients	EQ5D Index, EQ5D VAS, OHS, OKS	16-20 dependent on PROM model	23-30%
DoH (2012)	Retrospective cohort	Orthopaedic	NPROMS	2009-2010, 39,404 hip, 45,773 knees, 2010-11, 44,687 hips, 50,719 knees, 2011-12 47,392 hips, 54,062 knees.	EQ5D Index, EQ5D VAS, OHS, OKS	13-15 dependent on tool (some variable items listed & coded separately)	

Gutacker (2012)	Retrospective cohort	Orthopaedic	NPROMS	24,568 (2009-10 complete HES data)	EQ5D	7	
NHS England (2013a)	Retrospective cohort	Orthopaedic	NPROMS	As for DoH (2012)	EQ5D Index, EQ5D VAS, OHS, OKS	12 (some variable items listed & coded separately)	
Nuttall (2015)	Retrospective cohort	Orthopaedic	NPROMS	30,555 (2009-10 knee data post cleaning)	OKS	10 (some variable items listed & coded separately)	26% (OLS and FE model)

BM; baseline model, FOTO; Focus on Therapeutic Outcomes, FS; Functional Status, , FE; fixed effects, NPROMS; National Patient Reported Outcome Measures Programme, OKS; Oxford Knee Score, OHS; Oxford Hip Score, OHSM; Overall health status measure, OLS; ordinary least squares, PM; psychological measure, SF-12; Short Form 12, SF-36; Short Form 36.

Table 2-5 Summary of case-mix adjustment model variables

	Baseline PROM score	Age	Gender	Comorbidities	Duration of symptoms	Surgical history	Payer	Impairment type/procedure	Index of Multiple Deprivation	Exercise History	Ethnicity	Assistance with questionnaire	Disability	Living alone	Fear Avoidance Beliefs Questionnaire	Use of medication
First Author/s US																
Resnik 2003	x*	X	x		x*	x	x*			x						
Hart and Connolly 2006	x*	x*	x		x*	x	x	x		x						x
Hart 2011a	x*	x	x	x*	x*	x	x								x	
Hart 2011b	x*	x	x	x*	x	x*	x*			x					x	x
Resnik 2011	x	x	x	x	x		x			x						
Yen 2015	x	x	x	x	x	x	x									
Gozalo 2016	x*	x	x	x*	x*	x	x*	x							x	
Werneke 2016	x*	x*	x	x*	x*	x*	x*			x						x
First Author/s UK																
Browne 2007	x*	x*	x	x*	x	x*			x*							
Coles 2010	x*	x	x	x*	x	x		x	x*		x	x	x*	x		
DoH 2012	x*	x	x	x	x	x		x	x*		x	x*	x*	x		
NHS England 2013a	x*	x*	x*	x*	x			x*	x*		x*	x*	x*	x		
Gutacker 2012	x	x	x*	x*		x*		x*	x*							

Nuttall 2015	x	x	x	x	x	x	x	x	x	x		

Note: only variables used in 3 or more studies are included, * marks those identified in studies as most predictive variables

Resnik et al (2003) * 3 largest predictors

Hart and Connolly (2006) * 3 largest predictors

Hart et al (2011a) * 3 largest predictors

Hart et al (2011b) * 4 largest predictors

Resnik et al (2011) baseline model

Yen et al (2015) baseline model (all variables predictive)

Gozalo et al (2016) * 4 largest predictors

Werneke et al (2016) * 6 significant 'patient factor' predictors (retained in model)

Browne et al (2007) * 5 largest predictors (not including GH)

Coles (2010) * 4 largest predictors across models (not including GH)

DoH (2012) * 4 most predictive across models (not including depression)

NHS England (2013a) * 9 variables retained across primary hip/knee models

Gutacker et al (2012) * 5 largest predictors

Nuttall et al (2015) 10 significant variables included in model (not including length of stay)

MODEL DEVELOPMENT HISTORY:

US Model: Early FOTO models case-mix adjusted for 12 baseline variables as demonstrated by Hart and Connolly (2006) (see **Table 2-4**). All twelve variables were found to have a significant effect on discharge functional status score (FS) and predicted 35% of total variance, meaning that 35% of the variance in post treatment outcome could be explained by the model. The three independent variables with the largest partial R^2 values in the complete model were; intake FS, age, and symptom duration (Hart and Connolly, 2006), supporting the earlier paper by Resnik and Hart (2003). FOTO Inc. later moved to case-mix adjusting for eight patient variables, aware of the need to balance model power with feasibility of data collection (Hart and Connolly, 2006), as demonstrated in the paper by Hart et al (2011a), who looked at the effect of adding the variable of fear avoidance beliefs (FABQ-PA). Results demonstrated R^2 values of 0.2997 and 0.3010 respectively, with and without the inclusion of the FABQ-PA, thus improving model predictive ability but only slightly, and therefore this variable was not encouraged for future routine inclusion.

UK Model: In 2007 Browne et al (2007), set out to determine the feasibility of collecting pre- and post-operative PROMs data from patients undergoing elective surgery, and to develop methods to analyse and present collective data. Elective surgeries included five areas, with two MSK surgeries of interest: unilateral hip replacement and unilateral knee replacement. Significant variables within adjustment models were baseline PROM score, comorbidities, general health (GH), surgical history, age, and Index of Multiple Deprivation (IMD). Models explained between 24% and 27% of total variance.

Following the feasibility work by Browne et al (2007), Coles (2010) published the full UK NPROMS case-mix adjustment methodology (see **Table 2-5** for list of variables). Coles (2010) describes six orthopaedic models (separate models for each PROM, for each intervention). Variable numbers ranged from 16-20, and the models explained between 23% and 30% of total variance. All models

found the baseline score to be highly predictive of outcome, alongside patients who did not consider themselves disabled (positive impact), the GH question was highly predictive, and IMD and comorbidities were also predictive across models.

In 2011 increased data was available from the NPROMS collection which aided further model refinement, including changing the variables relating to co-morbidities and then removing the GH variable (DoH, 2012). Key predictive variables across updated models were baseline PROM score, disability status, comorbidity of depression, assistance with questionnaire, and IMD (DoH, 2012). In 2013, an alternative aggregation model (AAM) was proposed by NHS England (NHSE, 2013b) to further improve model stability. The full model was also updated following the separation out of primary and revision surgery (giving less prediction error). Significant changes to the model included the previous surgery variable being removed due to no longer being relevant, and inclusion of additional patient diagnostic codes. Key variables predicting outcome across updated primary hip and knee models were; baseline PROM score, age, sex, assistance with questionnaire, disability status, comorbidities, ethnicity, diagnostic codes, and IMD (NHSE, 2013a).

MODEL VALIDATION:

US Model: Hart and Connolly (2006) used two methods to validate the FOTO case-mix adjustment model. The patient sample was split into two, one to develop the model and one to test the stability of independent variables. 95% confidence intervals for the beta coefficients for all case-mix adjustment variables were similar. In the testing sample the predicted discharge FS was very close to the actual discharge FS (average predictive ratio 1.045) supporting model predictive validity, although the model slightly over predicted FS in the second sample. The paper by Hart et al (2011a) also carried out a split-half validation method to create a developmental and testing sample. No differences were found between beta coefficients between developmental and testing samples ($p<0.05$), again suggesting stability within the predictive model (Hart et al, 2011a).

UK Model: The inception NPROMS paper (Coles et al, 2010) considered the face validity of the case-mix adjustment models, appropriateness of scale, and direction and stability of the coefficients. The developed model was then tested in a subset of data. Comparisons between datasets and early testing suggested scope for removal of further variables either due to low incidence or volatility. The Knee EQ5D VAS model showed the only significant difference in samples due to the low incidence of some comorbidities, and lack of specific admission and discharge data. All models contained variables that appeared to be appropriate with directionally expected coefficients. Nuttall et al (2015) independently reviewed case-mix adjustment of NPROMS data. A comparison of mean predicted post-operative scores with mean actual scores using three statistical methods (ordinary least squares (OLS), fixed effects (FE), and random effects (RE) models), showed a fixed effects (FE) model performed the best (taking into account bias and precision), with a correlation coefficient of 0.800 when applied outside of the developmental sample (0.920 within the developmental sample) demonstrating high predictive ability (Nuttall et al, 2015).

MODEL STATISTICAL METHODS:

Statistical methods are used to explore the effect of individual and multiple explanatory variables on clinical outcome. The majority of studies used a stepwise approach when building a new regression model in order to make the most parsimonious model for clinical practice and used specific significance levels (0.05 (Hart and Connolly, 2006), 0.1 (Werneke et al, 2016) and 0.15 (Coles, 2010)) for inclusion/exclusion of independent explanatory variables. Early US and UK models used an ordinary least squares (OLS) multivariate regression method to estimate model power (R^2) (Hart and Connolly, 2006; Coles, 2010). Hierarchical models were demonstrated in later papers (Gomes et al, 2014; Yen et al, 2015; Gozalo et al, 2016). UK NPROMS moved to the use of a generalised least squares (GLS) method in 2011 (DoH, 2012). Support is growing for the use of GLS (NHSE, 2012; Nuttall et al, 2015) and hierarchical mixed models (Yen et al, 2015) that

take into account the nature and distribution of the data, including random clinic effects such as clustering (unmeasured factors within clinics that may affect outcome). The majority of latter papers therefore include using a stepwise approach to model development, and a GLS or hierarchical model for statistical analysis.

MODEL PREDICTIVE ABILITIES:

Using regression analysis, goodness of fit can be found by calculating R^2 which is usually expressed as a percentage. It explains the percentage of the variation in the dependent variable (PROM score) that can be explained by its relationship with the independent variables (patient factors) (Petrie and Sabin, 2013). Predictive ability across US study models ranged from 18-42% (Resnik and Hart, 2003; Hart and Connolly, 2006, Hart et al, 2011a; Hart et al, 2011b; Resnik et al, 2011, Yen et al, 2015; Werneke et al, 2016) and in UK models from 23-30% (Browne et al, 2007; Coles, 2010; Nuttall et al, 2015), meaning that between 18-42% of the post treatment PROM score can be explained by the included patient factor variables within models, demonstrating moderate to strong predictive ability across models (Cohen, 1988).

Discussion:

Table 2-5 details the patient factor variables used most commonly (those used in 3 or more studies) in included case-mix adjustment models. It can be seen that the most predictive and widely used variables across models include baseline PROM score, comorbidities, surgical history, IMD, age, payer, symptom duration, impairment type, assistance with questionnaire, self-reported disability, gender, and ethnicity. All of these variables are feasible for widespread clinical collection and warrant being considered for inclusion in future MSK case-mix adjustment modelling. Variables such as exercise history, living alone, FABQ, use of medication, and pain intensity had more limited support and therefore require further investigation before their inclusion can be fully justified. All US studies used the payer variable and all UK studies used the

IMD, these two variables may measure a similar construct as payer types have been used as proxy measures for a variety of demographic factors (Burstin et al, 1992; Yen et al, 2015).

Although there is considerable crossover in variables included within models, there is wide disparity in how variables are collated and entered into regression models, with a mixture of continuous, categorical and binary data. Models also used different PROM tools and different methods of collection. This would need to be considered when looking to test, replicate or build upon existing case-mix adjustment models, as when and how predictors are measured can have significant effects on model predictive performance (Whittle et al, 2017).

LIMITATIONS OF THE REVIEW:

The review included all studies focusing on the case-mix adjustment of MSK PROM data. The PROM used within studies was not limited. Studies can therefore not be fully compared with regards to statistical predictive ability, due to significant differences in PROMs utilised. The UK NPROMS methodology demonstrates that different variables are necessary dependent on which PROM tool is used (Coles, 2010, NHSE, 2013a). The review also included all healthcare settings including primary, community and secondary care. The limitation of this breadth is again the comparability of included studies, as patients across settings vary significantly in treatments received and outcomes realised. The review was also limited to those studies reported in the English language meaning that there may be models reported in languages other than English that have not been included. The majority of studies used convenience samples of complete datasets in healthcare data registries, which is the leading source of bias identified across studies. Results of this review therefore need to be viewed with caution until more robust prospective studies are undertaken.

SUMMARY OF FINDINGS:

Two broad case-mix adjustment models have been identified within the review. Neither model however has been externally validated. The two models are distinct in that one model is currently used within a community setting in the US (FOTO), and the other in a UK secondary care surgical setting (NPROMS). Future research is needed to externally validate existing models within and across MSK settings and countries, in order to be able to implement findings across healthcare settings and systems for the purposes of evaluating and improving patient care.

Recommendations for future MSK case-mix adjustment modelling of patient reported outcomes based on the combined findings within existing models are:

1. Patient factor variables warranting strong consideration for inclusion are: baseline PROM score, age, gender, comorbidities, symptom duration, surgical history, payer, impairment type, IMD, ethnicity, assistance with questionnaire, and self-reported disability.
2. A stepwise approach to model development is recommended, with significance levels of 0.05-0.15 demonstrated within included studies (Hart and Connolly, 2006; Coles, 2010; Werneke et al, 2016).
3. Statistical methods for consideration include GLS and hierarchical modelling which may be preferential to an OLS method due to accounting for clustering.
4. Methods need to minimise or account for missing data using structured prospective data collection and statistical methods such as data imputation or inverse probability weighting.
5. Defined PROM data capture at the start and end of treatment with a standardised follow up time-point is recommended to reduce risk of bias.

Conclusion:

Results demonstrate that there is strong evidence to support the use of case-mix adjustment modelling in MSK practice, and results highlight common areas of overlap between US and UK models, and models used within a community and secondary care setting. These results have

been summarised to aid development of case-mix adjustment methodology alongside much needed external validation of existing models, with the aim of optimising case-mix adjustment of MSK health outcomes. This will allow for effective performance profiling and future benchmarking of MSK services, both nationally and internationally.

3 Chapter 3: Identification of patient factor variables for use in case-mix adjustment of musculoskeletal PROM data.

The aims of this chapter were to; 1) carry out an umbrella review to identify predictors of MSK functional outcome, 2) explore the findings from **Chapter 2** in more depth with regards to potential case-mix adjustment model variables alongside any additional predictors identified in the umbrella review, to derive strength of evidence supporting inclusion of identified patient factor variables within future MSK case-mix adjustment modelling.

Objective 1: Predictors of functional outcome in musculoskeletal healthcare: An umbrella review

Abstract

Background: Multiple cohort and systematic review studies exist, reporting independent predictive factors associated with outcome in MSK populations. These studies have found evidence for a number of 'generic' factors that have been shown to predict outcome across MSK patient cohorts. This review provides a higher level review of the evidence with a focus on generic patient factors associated with functional MSK outcome with a view to informing predictive modelling.

Objectives: a) Identify patient factors found to have evidence to support their association with functional outcome, and b) review these findings across body areas/conditions to identify generic predictive factors.

Databases and Data Treatment: Electronic databases of MEDLINE, AMED, EMBASE, CINAHL and Cochrane were searched for eligible studies. Two reviewers independently extracted data and assessed quality using an established checklist for umbrella reviews.

Results: Twenty one systematic reviews met inclusion criteria, all were of moderate/high quality. Six independent predictors were found to have strong evidence of association with worse MSK functional outcome across anatomical body sites (worse baseline function, higher symptom/pain

severity, worse mental wellbeing, more comorbidities, older age and higher body mass index). Longer duration of symptoms, worse pain coping, presence of workers compensation, lower vitality and lower education were also found to have moderate evidence of association with worse functional outcome across body sites.

Conclusions: This study identifies a number of factors associated with musculoskeletal functional outcome. The generic predictive factors identified should be considered for inclusion into MSK prognostic models, including models used for case-mix-adjustment of patient reported outcome measure data.

Contribution to the Thesis

- This review identifies 'generic' patient factors that predict functional outcome (measured using PROMs) across MSK conditions.
- Findings provide support for the development and content of generic MSK prognostic models including models used to case-mix adjust PROM data for baseline complexity.
- When added to the findings from **Chapter 2** these findings allow for recommendations to be made on which variables to include within future MSK case-mix adjustment model development.
- Generic MSK models and functional PROMs would facilitate more feasible comparison and benchmarking of MSK services in order to identify variation and address health inequalities.

The contents of this review have been published in part in:

Burgess, R., Mansell, G., Bishop, A., Lewis, M. and Hill, J., 2020. Predictors of functional outcome in musculoskeletal healthcare: An umbrella review. *European Journal of Pain*, 24(1), pp.51-70.

Introduction

Patient Reported Outcome Measures (PROMs) consist of a series of questions that patients are asked in order to gauge their views on their own health, forming a self-assessment of a patient's health and health related quality of life (Devlin et al, 2010). Momentum among policy makers is growing for the routine and mandated collection and reporting of PROM data by clinical services. Outcomes interpretation however needs to give appropriate consideration to extraneous variables that can significantly impact upon treatment outcomes in patients with MSK impairments (Werneke et al 2016).

Case-mix adjustment is a way of statistically compensating for inter-provider differences in the prevalence of factors that adversely affect treatment response in order to make between-provider comparisons more equitable (Phillips et al, 2003). The use of adjustment methods assumes that the most important prognostic factors of treatment response have been identified and are appropriately measured and adjusted for (Deeks, et al, 2003). **Chapter 2** identified two broad MSK models and twelve baseline variables that were commonly used across studies and found to be predictive of functional outcome (Burgess et al, 2018). There was however, no formal review of key predictive factors in the development of either of the existing models. Important predictors may therefore have been omitted or identified since existing case-mix adjustment models were established.

Thousands of prognostic factor research studies are published within the medical literature each year (Riley et al, 2013, Altman and Riley, 2005). A large number of these cohort studies have focused on identifying predictors of outcome within a MSK context, including for specific MSK conditions such as low back pain, shoulder pain, and hip pain, with well conducted systematic reviews summarising agreement on the most important predictive factors (Verkerk et al, 2012, Kooijman et al, 2015, Struyf et al, 2016, de Rooij et al, 2016). Predictive patient factors emerging from these reviews include; symptom duration, disability level, previous episode, pain severity,

baseline function, and level of comorbidity (Kooijman et al, 2015, Struyf et al, 2016, de Rooij et al, 2016). Mallen et al, (2007) proposed that many of these predictors were common across pain sites and conducted a systematic review to identify prognostic factors for generic MSK pain. They found that although there was a high level of heterogeneity across studies, certain generic patient factors were consistently found to be predictive of poor outcome. This review was updated in 2017 (Artus et al, 2017) and generic predictors of MSK outcome were found to be; widespread pain, high functional disability, somatisation, high pain intensity and presence of previous pain episodes. These studies support the notion that there are key generic factors across pain sites/MSK conditions that predict treatment outcome.

Umbrella reviews are reviews of existing systematic reviews (Aromataris et al, 2015). Their purpose is to summarise evidence from multiple research syntheses (Becker et al, 2011) and to provide a rapid and broad review of the evidence base within a specified field (Khangura et al, 2012, Lunny et al, 2017). They are increasingly conducted due to a steady increase in the number of systematic reviews undertaken. This allows for additional analysis in comparing and contrasting systematic review findings, providing a synopsis of high-level research evidence (Aromataris et al, 2015).

The aim of this study was to conduct a high-level umbrella review of all existing systematic reviews providing longitudinal data on self-reported baseline predictors of functional outcome within MSK patient populations, with the purpose of providing robust evidence on prognostic factors that should be considered when developing MSK case-mix adjustment models for adjusting PROM scores to allow for "fair" comparison of data across healthcare services/providers.

Methods

Design: Umbrella review in line with guidelines from the umbrella review methodology working group (Aromataris et al, 2015).

OBJECTIVES

The objectives of this umbrella review were to; a) identify patient factors found to have evidence to support their predictive association with functional outcome, measured using an MSK-relevant PROM score in patients presenting with MSK conditions, and b) review these findings across body areas/conditions to identify generic predictive factors.

INCLUSION AND EXCLUSION CRITERIA

Specific inclusion and exclusion criteria were formulated for the review to ensure studies included were justified, relevant and reviewed against clear and consistent criteria (see **Table 3-1**). Inclusion criteria were; systematic reviews and meta-analyses, adult populations, English language (due to resource limitations), all MSK healthcare settings (primary care, community, secondary care, occupational), studies identifying independent patient level predictors of functional outcome in patients presenting with MSK disorders, self-reported predictors (to ensure feasibility of clinical collection), reviews had to include functional outcome as one of the primary outcomes, functional outcome measured using PROMs including those used to measure disease severity, disability and functional activity, studies published in the last five years ((01/01/2013-01/01/2018). Aromataris et al (2015) state that including research syntheses conducted within the past 5–10 years reflects original/primary research conducted over the past 30 years, and therefore restricting reviews to these time periods in this type of review is appropriate. Exclusion criteria were studies looking at predictive models rather than individual factors. Variables included needed to be gathered at baseline, self-reported, and outside of a provider's control, with the focus on identifying generic variables that could be feasibly collected in routine practice for use in case-mix adjustment predictive modelling. Prognostics factors are characteristics that help to estimate patient's likely outcome irrespective of chosen management

(Hill and Fritz, 2011), therefore all systematic reviews were included within an MSK patient population irrespective of management delivered.

Table 3-1 Inclusion and exclusion criteria

Inclusion Criteria	Exclusion Criteria	Justification
Systematic Review		Highest level of evidence
Adult Population		Focus to development of adult case-adjustment model
English Language		No availability of interpreter
All MSK healthcare settings (primary care, community, secondary care, occupational)		MSK focussed
Identifying independent predictors of outcome	Predictive models	Focus on additional candidate variables that could be added to current case-mix adjustment models
Predictors to include self-reported patient factor variables, that are feasible for clinical collection	Predictors not to include diagnostic, treatment classification, or service level variables.	Variables must be able to be feasibly gathered at baseline and be outside of a Providers control
Functional outcome measured using PROM to include those measuring disease severity, disability and functional activity.	No PROM used	Focus on identifying additional variables to adjust PROM data
Reviews had to include functional outcome as one of their primary outcomes	Functional outcome not a core focus of the review	Focus on functional outcome
Published in last 5 years 01/01/2013-01/01/2018 inclusive	Reviews published prior to 01/01/2013	Main focus on identification of additional variables beyond the development of existing risk-adjustment models. Searching past 5 years will still include primary research from past 30 years (Aromataris et al, 2015).

SEARCHES

Medline, AMED, EMBASE and CINAHL electronic databases were searched alongside the Cochrane Database of systematic reviews, for systematic reviews in the last 5 years (since the development of the latest iteration of the NPROMS model (NHS England, 2013), and likely to

encompass primary research from past 30 years (Aromataris et al, 2015)) (search strategy for Medline included within **Appendix 3-1**). Search criteria are listed in **Table 3-2**. Searches were combined using Boolean logic (AND and OR). Search criteria included having the following terms, or iterations of them, in the study title or abstract; predict OR prognosis, AND, outcome OR recovery OR function, AND, musculoskeletal OR low back OR neck OR spinal OR hip OR knee OR shoulder. Reference lists of relevant identified articles were also searched to identify any further appropriate reviews. Electronic searches were filtered to ensure results included systematic reviews only.

Table 3-2 Search terms

Type of study (prognostic)	Outcome (functional status)	MSK Condition
Predic* Prognos*	Outcome Recovery Function	Musculoskeletal (or) Low back Neck Spinal Hip Knee Shoulder

SELECTION PROCEDURE

One reviewer (RB) undertook a preliminary screen of all titles to remove studies that clearly and unquestionably did not meet the inclusion criteria. Two reviewers (RB and GM) then independently screened all abstracts identified from searches and accessed full articles for all abstracts supporting closer consideration for inclusion. The full text of identified papers was then reviewed independently by the same reviewers. Any disagreement on studies to include was discussed until consensus was reached. A third reviewer was also available if agreement could not be reached (JH).

DATA EXTRACTION AND QUALITY ASSESSMENT

Data was independently extracted by two reviewers (RB and GM) and entered into an excel table, data extracted included title, authors, year, date range of database searching, setting, MSK condition/body part, number of studies included and quality, quality tool used, outcomes reported, variables included, and strong, moderate, weak, or inconclusive evidence of variable predictive ability.

Quality was assessed using a checklist developed by the umbrella review methodology working group (Aromataris et al, 2014) (developed specifically for appraising systematic reviews within the context of an umbrella review) that was designed for use by two reviewers (RB and GM) independently appraising studies with discussion where necessary to reach agreement. The checklist includes 10 items, including for example; whether the review question was clearly and explicitly stated, whether the search strategy was appropriate, and whether the criteria for appraising studies was appropriate. The full checklist is shown in **Table 3-4** with agreement achieved on all studies.

Similar to scoring of other systematic reviews using AMSTAR (Shea et al, 2007), studies were graded as low quality if three or less items were rated as met (0-3), moderate quality (4-7), and high quality (8-10) (Rebar et al, 2015).

SUMMARISING EVIDENCE ACROSS STUDIES

Criteria for evaluating the level of evidence across studies was adapted from Sackett (2000) and from previous similar review studies (Kooijman et al (2015), Artus et al (2017)), and used to determine the strength of evidence for independent predictors of functional outcome reported where similar criteria had not already been applied (see **Table 3-3**). To be considered a generic predictor of functional outcome, prognostic factors had to be investigated for two or more musculoskeletal pain regions (Artus et al, 2017). Like Artus et al (2017), synthesis of results took

into consideration statistically significant associations, consistency of results (direction of effect), and study quality.

Table 3-3 Level of evidence for generic prognostic factors for poor functional outcome (Adapted from Sacket (2000), Kooijman et al (2015), Artus et al (2017))

Level of Evidence	
Strong	Consistent significant findings (≥75 % of studies) in high quality studies (at least 2)
Moderate	Consistent significant findings (≥75 % of studies) in high and low quality studies (at least 1 high quality study in the direction of consistent significant findings)
Weak	Significant findings in only 1 high quality study or consistent findings (≥75 % of studies) in at least three or more low quality studies
Inconclusive	Inconsistent findings irrespective of study quality, or less than three low quality studies available

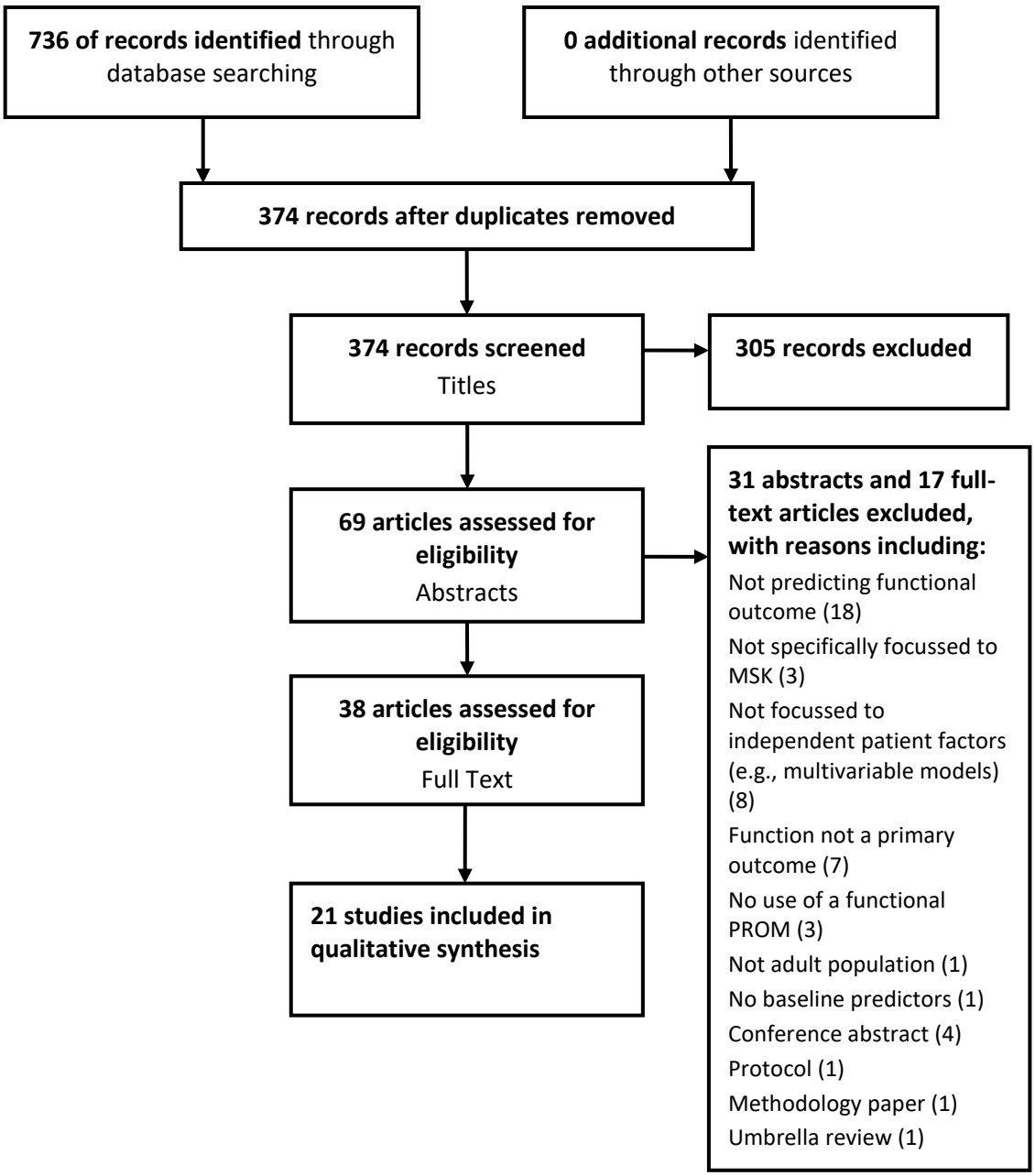
Guyatt et al (2008) describe the GRADE system in a series of articles which encompasses how to grade quality of evidence and the strength of recommendations formulated through evidence syntheses. This guideline on recommendations was also used to help inform and underpin the formulation of recommendations from the review.

Results

SEARCH RESULTS AND SELECTION PROCEDURE

The above searches identified 374 systematic review articles for initial screening after removal of duplications (no additional articles were found from searches of reference lists). Following review of titles, 69 reviews were included for closer inspection of abstracts, and 38 for full text review. Following full text review 21 studies were included. See **Figure 3-1** for flowchart of search results.

Figure 3-1 Flowchart of search results



QUALITY ASSESSMENT

The quality checklist is shown with results for included studies in **Table 3-4**. For each criterion within the checklist, studies were rated as met, not met, or not applicable (NA).

Table 3-4 Quality of included studies (Critical appraisal checklist for systematic reviews and research syntheses (Aromataris et al, 2014))

First Author	Quality Criteria										Quality Rating
	Review question clear and explicit	Inclusion criteria appropriate	Search strategy appropriate	Sources and resources used adequate	Appraisal criteria appropriate	Appraisal by 2 or more reviewers independently	Methods for combining appropriate	Likelihood of publication bias assessed	Recommendations for policy/practice supported	Specific directives for new research appropriate	
Bastick 2015	Met	Met	Not met	Not met	Met	Met	NA	Not met	Met	Not met	Moderate
Bruls 2015	Met	Met	Not met	Met	Met	Met	NA	Met	Met	Met	High
Buirs 2016	Met	Met	Not met	Met	Met	Not met	NA	Not met	Met	Met	Moderate
Chester 2013	Met	Met	Met	Met	Met	Met	NA	Not met	Met	Met	High
De Rooij 2016	Met	Met	Met	Met	Met	Met	NA	Met	Met	Met	High
Harmelink 2017	Met	Met	Met	Not met	Met	Met	NA	Not met	Met	Met	Moderate
Heerspink 2014	Met	Met	Met	Met	Met	Met	NA	Not met	Met	Met	High
Hofstede 2016	Met	Met	Met	Met	Not met	Met	NA	Not met	Met	Met	Moderate
Kooijman 2015	Met	Met	Met	Met	Met	Met	NA	Not met	Met	Met	High
Lungu 2016a	Met	Met	Met	Met	Met	Met	NA	Not met	Not met	Not met	Moderate
Lungu 2016b	Met	Met	Met	Met	Met	Met	NA	Met	Not met	Not met	Moderate
Magklara 2014	Met	Met	Met	Met	Not met	Met	NA	Not met	Met	Met	Moderate
McKillop 2014	Met	Met	Met	Met	Met	Met	NA	Not met	Not met	Not met	Moderate
Pinheiro 2016	Met	Met	Met	Met	Met	Met	NA	Not met	Met	Met	High
Struyf 2016	Met	Met	Met	Met	Met	Not met	NA	Not met	Met	Met	Moderate
Wertli 2014a	Met	Met	Not met	Met	Met	Met	Met	Not met	Met	Met	Moderate
Wertli 2014b	Met	Met	Not met	Met	Met	Met	NA	Not met	Met	Met	Moderate
Wertli 2014c	Met	Met	Not met	Met	Met	Met	NA	Not met	Met	Met	Moderate
Wertli 2014d	Met	Met	Not met	Met	Met	Met	NA	Met	Met	Met	High
Wilson 2016	Met	Met	Met	Met	Met	Met	NA	Not met	Met	Met	High
Woollard 2016	Met	Met	Met	Not met	Met	Not met	NA	Not met	Met	Met	Moderate

QUALITY OF INCLUDED STUDIES

Overall quality across included reviews was good (moderate to high quality) with the majority of checklist criteria rated as 'met' (See **Table 3-4**). The majority of reviews (13/21) stated following specific review guidelines for reporting review methodology and findings (PRISMA, MOOSE). All reviews had more than one independent reviewer apart from the studies by Struyf et al (2016), and Buirs et al (2016). All included systematic review studies used a quality tool to grade included studies with regards to quality of evidence, although Hofstede et al (2016) limited this to grading based on two criteria only. Most studies were unable to carry out a meta-analysis due to heterogeneity of included studies/data and therefore this item was scored NA (not applicable). Not all studies fully assessed likelihood of publication bias, and therefore this item was often unclear/not met.

The majority of reviews describe the main limitations across included studies to be: heterogeneity of methodology of primary studies, including a diverse range of outcome measures and prognostic factors across studies, differences in timing of outcome capture, incomplete data, and loss of patients to follow up.

Evidence Synthesis

Most included studies used criteria to determine the level of evidence for individual predictors such as those shown in **Table 3-3**, which grouped levels of evidence into strong, moderate, weak, or inconclusive. These criteria were utilised by the reviewers (RB, GM) to categorise predictors where this had not already been undertaken, with results shown in **Tables 3-5**, **3-6 and 3-7**. Where possible only functional outcome was reported unless this could not be separated from other primary outcomes within studies.

Table 3-5 Characteristics of included studies

First Author	Study Characteristics Body area/condition	Interventions/phenomena of interest	Setting	Measure used to determine functional outcome	Types of studies	Dates databases searched	Studies included
Bastick 2015	Knee OA	Longitudinal data	All MSK settings	WOMAC-PF	Prospective cohorts	Inception to 2015	30 (20 high quality)
Bruls 2015	Arm, neck, shoulder	Longitudinal data	All MSK settings	SDQ, DMQ, NDI, DASH SF-36 PF	Prospective cohorts	1966 to 2013	26 (16 high quality)
Buirs 2016	Hip OA	Total hip arthroplasty	Orthopaedic surgical	HHS, OHS, SF36, LEFS, WOMAC	Prospective cohorts	Inception to 2015	33 (5 of high quality)
Chester 2013	Shoulder pain	Physiotherapy intervention	All MSK settings	CMS, UCLA, ASES, SPADI, FS, DASH, FLEX-SF	Prospective cohorts or trial analysed as cohort	Inception to 2013	33
De Rooij 2016	Hip OA	Longitudinal data	All MSK settings	Not stated	Prospective cohort or trial analysed as cohorts	Inception to 2015	15 (11 high quality)
Harmelink 2017	Knee OA	Total knee arthroplasty	Orthopaedic surgical	WOMAC, OKS, IKSS, SF-12, KSS, SF-36, AKSS.	Prospective and retrospective cohorts	2000 to 2016	18
Heerspink 2014	Shoulder	Rotator cuff repair	Orthopaedic surgical	DASH, CMS, ASESSS	Prospective cohorts	1929 to 2013	12 (1 high quality)
Hofstede 2016	Hip OA	Total hip arthroplasty	Orthopaedic surgical	SF-36, EQ5D, SF-12, WOMAC, OHS, HHS	Prospective cohorts	Inception to 2014	35 (9 high quality)

Kooijman 2015	Neck and shoulder	Longitudinal data	All MSK settings	SPADI, DASH, UCLA	Prospective and retrospective cohorts	2003 to 2014	9 new articles (6 high quality) (25 including previous review)
Lungu 2016a	Hip OA	Total hip arthroplasty	Orthopaedic surgical	WOMAC, HHS, OHS, LEFS, HOOS	Prospective cohorts	Inception to 2015	22 (mean quality score 81% (moderate to high quality)), 4 scored >90%
Lungu 2016b	Knee OA	Total knee arthroplasty	Orthopaedic surgical	WOMAC, OKS	Prospective cohorts	Inception to 2014	33 (mean quality score 80.7% (moderate to high quality), 9 scored >90%
Magklara 2014	Knee OA	Total knee arthroplasty	Orthopaedic surgical	WOMAC, HHS, AKSS, SF-36, PFS, FLP	Prospective cohorts	Inception to 2013	8 (all good/high quality)
McKillop 2014	Low back pain	Lumbar spinal stenosis surgery	Orthopaedic surgical	SSSQ, ODI	Prospective cohorts	1980 to 2012	13 (only high-quality studies included)
Pinheiro 2016	Low back pain	Longitudinal data	All MSK settings	RMDQ, ODI	Prospective cohorts	Inception to 2014	17, 13 cohorts (1 high quality meeting all criteria, average score 70%)
Struyf 2016	Shoulder pain (non-traumatic)	Longitudinal data	All MSK settings	SPADI	Prospective cohorts	Inception to 2014	9 (7 high quality)
Wertli 2014a	Low back pain	Longitudinal data	All MSK settings	ODI, RMDQ, SF-36 PF	Prospective cohorts	1980 to 2012	19 publications, 16 cohorts. (4 high quality)
Wertli 2014b	Low back pain	Longitudinal data	All MSK settings	ODI, RMDQ, SF-36 PF	RCTs analysed as cohorts	1980 to 2012	13 publications, 11 RCTs (7 high quality)
Wertli 2014c	Low back pain	Longitudinal data	All MSK settings	ODI, RMDQ	Prospective cohorts	1990 to 2011	21 (4 high quality)

Wertli 2014d	Low back pain	Longitudinal data	All MSK settings	ODI, RMDQ	RCTs analysed as cohorts	1990 to 2013	18 publications, 17 RCTs (5 high quality)
Wilson 2016	Low back pain	Lumbar discectomy	Orthopaedic surgical	ODI, EQ5D, Sf-36, JOAS, RMDQ, ODI, SBI, SFI, PDS, NOS	RCTs, controlled trials or prospective cohorts	Inception to 2014	40 (all high quality)
Woollard 2016	Shoulder pain	Rotator cuff repair	Orthopaedic surgical	CMS, ASES, DASH	Prospective and retrospective cohorts	1995 to 2015	23 (1 study scoring 5/7 in quality assessment and 3 studies 4/7 indicating higher quality)

* Intervention/phenomena of interest was classed as 'longitudinal' if there was no specific intervention of interest. All studies however were longitudinal.

** All MSK settings included all MSK healthcare settings (primary care, secondary care, community, occupational)

AKSS; American Knee Society Score, ASES; American Shoulder and Elbow Surgeons Standardised Shoulder Assessment, CMS; Constant-Murley Score, DASH; Disabilities of the Arm and Shoulder, DMQ; Dutch Musculoskeletal Questionnaire, FIQ; Functional Index Questionnaire, FLP; Functional Limitations Profile, FLEX-SF; Flexilevel Scale of Shoulder Function, FS; Functional Status, HHS; Hip Harris Score, HOOS; Hip Disability and Osteoarthritis Outcome Score, IKSS; International Knee Society Score, JOAS; Japanese Orthopaedic Association Score, LEFS; Lower Extremity Functional Scale, LSS; Lumbar Spinal Stenosis, NDI; Neck Disability Index, NOS; Newcastle Ottawa Scale, NPOS; Neck Pain Outcome Score, OA; Osteoarthritis, OHS; Oxford Hip Score, PDS; Pain Disability Score PF; Physical Function, PFS; Physical Functioning Scale, PFJ; Patellofemoral Joint, PRWE; Patient Related Wrist Evaluation, PFP; Patello-Femoral Pain, RCR; Rotator Cuff Repair, RMDQ; Roland Morris Disability Questionnaire, SBI; Sciatica Bothersome Index, SDQ; Shoulder Disability Questionnaire, SFI; Sciatic Frequency Chart, SPADI; Shoulder Pain and Disability Index, SST; Simple Shoulder Test, SSSQ; Swiss Spinal Stenosis Questionnaire, TKA; Total Knee Arthroplasty, THA; Total Hip Arthroplasty, WOMAC-PF; Western Ontario and McMaster Universities Osteoarthritis Index Physical Function, WORC; Western Ontario Rotator Cuff Index, WOSI; The Western Ontario Shoulder Instability Index, UCLA; UCLA Shoulder Score.

Table 3-6 Studies showing evidence for predictive factors associated with functional outcome

Authors	Predictors of Poor Outcome			Predictors of Good Outcome			Inconclusive
Evidence Level	**Strong**	**Moderate**	**Weak**	**Strong**	**Moderate**	**Weak**	**Inconclusive**
Bastick 2015	Age, ethnicity, BMI, Comorbidity count.	Educational level, vitality, pain coping subscale resting	Pain coping subscales worrying, hoping and catastrophising; knee injury; knee surgery; bisphosphonate usage				Gender, mental health, bisphosphonate usage, bodyweight change
Bruls 2015	Baseline function (ST), coping (ST), presumed cause (ST).		Job stress (ST), catastrophising (LT).				Age, paid work (ST), children, symptom duration, comorbidities, past trauma (ST), symptom severity (ST), ergonomic risk factors, general health, catastrophising (ST), social class, baseline function (LT), coping (LT).
Buirs 2016	BMI, age, pre operative physical function, greater comorbidity.		Education	Better mental health			Gender, socioeconomic status, alcohol consumption, allergies, vitamin D insufficiency.

Chester 2013	Higher disability, longer duration of symptoms		Increasing age				
De Rooij 2016	Higher comorbidity count, lower vitality (SF36)		Moderate or severe cardiac or ENT disease, presence of CIRS, poor GH perception, hip morning stiffness <60 mins, bilateral hip pain with equal symptoms, reduced hip flexion at baseline, presence of knee OA, bilateral knee pain, knee morning stiffness <30 mins, reduced knee extension baseline, no supervised exercise, lower level physical activity, high bodily pain, avoidance activity.				Lower level education, more disability, BMI, higher hip pain at baseline, poor cognitive functioning, resting, transformation.
Harmelink 2017						Lower pre-operative function (higher change)	Age, absence of anxiety, presence of social support,

							higher income, normal BMI, less comorbidity, gender
Heerspink 2016		Workers compensation board status	Additional AC/biceps surgical procedure				Age, smoking, traumatic onset, symptom duration, obesity, comorbidity, preoperative expectations
Hofstede 2016	Preoperative function (lower score predicts greater improvement but worse outcome), Worse mental wellbeing					Higher education/socioeconomic status	Comorbidities, BMI, pre-operative pain, gender, age, expectations, QoL
Kooijman 2015	Primary Care: Higher shoulder pain intensity, concomitant neck pain, longer duration of symptoms. Secondary Care: greater disability.			Secondary care: no previous shoulder pain	Secondary care: higher education		Primary care: greater disability, previous episode of pain, gender, gradual onset. Secondary care: gradual onset, long duration of complaints, diagnosis, physical workload, no previous shoulder pain, non-dominant side, health status. Occupational setting; longer

							duration of symptoms, higher age, work related psychosocial factors, high physical workload, female gender
Lungu 2016a THA	Pre-operative functional status (lower associated with lower post op score but greater change), higher BMI, higher comorbidity, worse general/mental health (SF36/SF12)		Lower education				Age, living alone, expectations, widespread pain.
Lungu 2016b TKA	Pre-operative functional status (lower associated with lower post op score but greater change).	Presence of back pain, pain catastrophizing, pre-operative mental/general health (SF36)	BMI				Age, gender, socioeconomic status, depression/anxiety, comorbidities.
Magklara 2014							Self-efficacy
McKillop 2014	Depression (predicted greater disability and symptom severity)						

Pinheiro 2016	Depression (predicted greater disability)						
Struyf 2016	Duration of symptoms, baseline pain score, baseline disability score	Gender (male), Age, GP visits, Sick leave duration, Poor general health, Gradual onset, Perceived job demand, Perceived social support.			Not regular medication, Active treatment		Education, shoulder dominance, locus of control, previous shoulder pain, previous neck pain, other diseases, concomitant neck pain, concomitant psychological complaints, Causes (all), Job: shoulder movements per minute, repetitive movements, perceived job control, use of shoulder force, overhead work, task cycle duration, Psychosocial factors (all)
Wertli 2014a (catastrophising, observational studies only)							Catastrophising
Wertli 2014b (catastrophising, RCTs only)		Greater catastrophising					

Wertli 2014c (fear avoidance, observational studies only)							Fear avoidance beliefs
Wertli 2014d (fear avoidance RCTs only)			Fear avoidance (ST outcome <6 months)				
Wilson 2016	Symptom severity, Workers compensation (long sick leave time)	Reoperation, Workers compensation (compensation)	Comorbidities, Socioeconomic status, Expectations, Anxiety, Pre-operative ODI, Joint pain, Workers compensation (restricted duties)	Age (younger), Better mental health, More severe leg pain, Absence of workers compensation	Symptom duration	Gender (male), Expectations, Mental Health ((SF-36) better), Pain duration (less), Pain frequency (less), Pain severity (less) (SF-36 body pain), Pain severity (high) (Back pain VAS)	Smoking, gender (female), obesity, age, expectations, depression, pain dominance, duration of leg pain, work type.
Woollard 2016							

AC; Acromioclavicular; BMI; Body Mass Index, LT; Long Term, ODI; Oswestry Disability Index, QoL; Quality of Life, RCT; Randomised Controlled Trial, ST; Short Term, VAS; Visual Analogue Scale

Table 3-7 Generic predictors of functional outcome

Category	Predictor of functional outcome	Shoulder	Hip	Knee	Spine	Strength of outcome prediction	Generic Predictor
Baseline Function	**Poor Outcome:** Baseline Function/Disability (worse)	Bruls, 2015 (strong), Kooijman, 2015 (strong), Struyf, 2016 (strong), Chester 2013 (strong)	Hofstede, 2016 (strong), Buirs 2016 (strong), Lungu, 2016a (strong)	Lungu, 2016b (strong)	Wilson, 2016 (weak)	Strong	**Yes (Strong)**
	Good Outcome: Baseline Function/Disability (worse)			Harmelink, 2017 (weak) (lower function higher change)		Inconsistent with poor outcome but includes change scores	
Baseline Symptom Severity	**Poor Outcome:** Baseline Pain Intensity/Symptom Severity (higher)	Kooijman, 2015 (strong), Struyf, 2016 (strong)			Wilson, 2016 (symptom severity) (strong)	Strong and consistent across poor outcome	**Yes (Strong)**
	Good Outcome: Baseline Pain Intensity/Symptom Severity (lower)				Wilson, 2016 (back pain VAS high) (weak), Wilson, 2016 (leg pain severity high) (strong)	Inconsistent with poor outcome	
Mental Wellbeing	**Poor Outcome:** Mental Wellbeing/Depression/Anxiety (worse)		Hofstede, 2016 (strong)		McKillop, 2014 (strong), Pinheiro, 2016 (strong), Wilson, 2016 (weak).	Strong	**Yes (Strong)**

	Good Outcome: Mental Wellbeing/Depression/Anxiety (less/absent)		Buirs, 2016 (strong)	Bastick, 2015 (weak)	Wilson, 2016 (strong)	Consistent with poor outcome providing strong evidence	
Comorbidities	**Poor Outcome:** Comorbidities (more)		Buirs 2016, (strong), de Rooij, 2016 (strong), Hofstede, 2016 (weak),	Lungu, 2016b (strong), Bastick, 2015 (strong)	Wilson, 2016 (weak)	Strong	**Yes (Strong)**
	Good Outcome: Comorbidities (less)						
Age	**Poor Outcome**: Age (older)	Chester, 2013 (weak), Struyf, 2016 (weak)	Buirs, 2016 (strong)	Bastick, 2015 (strong)		Strong	**Yes (Strong)**
	Good Outcome: Age (younger)				Wilson, 2016 (strong)	Consistent with poor outcome providing strong evidence	
BMI	**Poor Outcome:** Higher BMI		Buirs, 2016 (strong), Lungu, 2016a (weak)	Bastick, 2015 (strong), Lungu, 2016b (strong)		Strong	**Yes (Strong)**
	Good Outcome: Lower/normal BMI						
Symptom Duration	**Poor Outcome:** Duration of symptoms (higher)	Chester, 2013 (strong), Kooijman, 2015 (strong), Struyf, 2016 (strong)				Strong but only in 1 body area	**Yes (Moderate)**

	Good Outcome: Duration of Symptoms (lower)				Wilson, 2016 (moderate)	Consistent with poor outcome providing moderate evidence	
Pain Coping	**Poor Outcome:** fear avoidance/catastrophising (high)	Bruls, 2015 (weak)		Bastick 2015 (moderate), Lungu, 2016b (moderate)	Wertli 2014b, (moderate), Wertli 2014d (weak)	Moderate	**Yes (Moderate)**
	Good Outcome: fear avoidance/catastrophising (low)						
Workers Compensation	**Poor Outcome:** Workers compensation/Sick leave duration (Present/longer)	Heerspink, 2014 (moderate), Struyf, 2016 (weak)			Wilson, 2016 (strong)	Moderate	**Yes (Moderate)**
	Good Outcome: Workers Compensation/Sick leave duration (absent/less)				Wilson, 2016 (strong)	Consistent with poor outcome providing moderate evidence	
Vitality	**Poor Outcome** (lower)		de Rooij, 2016 (strong)	Bastick 2015 (moderate)		Moderate	**Yes (Moderate)**
	Good Outcome (higher)						
Education	**Poor Outcome** (lower education/socioeconomic status)		Lungu, 2016a (strong)	Bastick, 2015 (moderate)		Moderate	**Yes (Moderate)**

	Good Outcome (higher education/socioeconomic status))	Kooijman, 2015 (moderate)	Hofstede, 2016 (weak)			Consistent with poor outcome providing moderate evidence	
General Health	**Poor Outcome:** Poor general/mental health		Lungu, 2016a (moderate)	Lungu, 2016b (strong)		Moderate	**Yes (Moderate)**
	Good Outcome: Better general/mental health						
Widespread pain	**Poor Outcome:** Widespread/body pain (greater)		Lungu, 2016a (weak), de Rooij, 2016 (weak)	Lungu, 2016b (weak)		Weak	**Yes (Weak)**
	Good Outcome: Widespread/body pain (lower)				Wilson, 2016 (weak)	Consistent with Poor Outcome providing weak evidence	

PREDICTORS OF OUTCOME

Predictors of outcome including those with strong and high-quality consistent evidence, moderate evidence, weak evidence, and inconclusive or inconsistent/conflicting evidence in predicting functional outcome are shown in **Table 3-6.** Those variables with; strong, moderate, or weak evidence of effect on functional outcome that could be considered generic predictors (across two or more studies with different MSK pain sites/conditions) are shown in **Table 3-7**. Similar predictors are grouped together where appropriate. **Table 3-7** also shows where there are any inconsistencies in regards to the direction of effect within the column termed 'strength of outcome prediction'. The table shows that there are a number of predictors that could be considered generic predictors of functional outcome, and these are discussed with consideration to the strength of evidence below.

Strong evidence: Variables with a strong level of evidence supporting their ability to predict functional outcome across more than one body area are; baseline function/disability, baseline symptom/pain severity, mental wellbeing, comorbidities, age and body mass index (BMI). The majority of studies show that poor outcome is predicted by; worse baseline function/disability, higher pain severity, worse mental wellbeing, more comorbidities, older age, and higher BMI (see **Table 3-7** for detail of inconsistencies in direction of effect)

Moderate evidence: Variables with a moderate level of evidence supporting their ability to predict functional outcome across more than one body area are; duration of symptoms, pain coping (including fear avoidance beliefs and catastrophizing), workers compensation/sick leave duration, vitality (SF36), education/socioeconomic status and general health. Poor outcome was consistently predicted by; longer duration of symptoms, higher levels of fear avoidance/catastrophizing, the presence of workers compensation or higher sick leave duration, lower vitality, lower education, and poor general health.

Weak evidence: The variable of widespread pain was found to have weak evidence across more than one body area/condition, with greater widespread pain consistently predicting poor outcome.

Inconclusive evidence: A large number of variables were shown to have inconclusive evidence, see **Table 3-6**.

Discussion

In summary the main independent factors with strong evidence for predicting functional outcome in MSK patient cohorts across body sites from this umbrella review were worse baseline function/disability, higher symptom/pain severity, worse mental wellbeing, more comorbidities, older age, and higher BMI. Longer duration of symptoms, poorer pain coping, presence of workers compensation/sick leave duration, lower vitality, lower education /socioeconomic status, and poorer general health were also found to have moderate evidence to support them being generic predictors of functional outcome, and presence of widespread pain weak evidence.

The majority of studies showed consistent findings with regards to the direction of effect (see **Table 3-7**) but there were two specific areas of inconsistency. The first was in relation to baseline function/disability with the majority of studies showing that worse baseline function predicted poor outcome, this was however not the case when studies looked at change scores rather than purely follow up outcome, for example Hofstede et al (2016) found that a lower functional PROM score (indicating lower function) predicted worse outcome conforming to other studies, but also to greater improvement (change). This was also found by Harmelink et al (2017), and Lungu et al (2016a, 2016b). This shows that patients with lower functional scores achieve a better outcome with regards to health gain but still a worse overall outcome than those with higher baseline scores in relation to follow up functional PROM scores achieved. All of these studies were looking at predictors of surgical hip or knee arthroplasty functional outcome so it also unclear if this phenomenon is specific to this type of intervention. There is therefore clear evidence that

baseline functional PROM score is predictive of outcome across included studies, but this needs to be viewed in context to how it is being evaluated. The second area where there was a degree of inconsistency was in evaluating baseline pain/symptom severity. The majority of studies showed worse baseline pain/symptom severity to predict worse outcome but this was less consistent in the area of spinal surgery where worse baseline symptoms predicted worse outcome but specifically more severe leg pain predicted better post-operative outcome. This is likely due to leg pain being a treatment effect modifier within the area of lumbar discectomy surgery, meaning that this subgroup of patients respond differently/better to this specific treatment (Hancock et al, 2009). Within this review we were looking to determine predictors of functional outcome across MSK patient populations to allow for a more evidence-based approach to adjustment of case-mix to take into consideration those at a higher or lower risk of a poor outcome, not to provide information on which patients respond better to which specific treatments (modifiers of treatment outcome) (Hancock et al, 2009).

This review of independent predictors of MSK functional outcome largely reflects the review findings of Artus et al (2017) and of Mallen et al (2007) on prognostic factors for patients with MSK pain. Artus et al (2017) found generic predictors of a poor prognosis with strong evidence to be; widespread pain, high functional disability, and somatisation, and predictors with a moderate level of evidence to be; high pain intensity, long pain duration, and a high depression/anxiety score. Evidence for no association was also found for; low education (strong), pain medications (moderate), and older age and gender (weak). This demonstrates that, whether focused to prognostic factors for MSK pain or specifically to prognostic factors for functional outcome in MSK patients, generic factors are largely similar. This review does however provide different findings with regards to education and age and provides evidence for additional factors of BMI, vitality, workers compensation, comorbidities, and general health. From this study it can be seen that factors such as BMI and vitality predict functional outcome in patients with clinical OA but it

is unclear whether this would be the case for other MSK pain conditions as this factor was not evaluated outside of this patient population although it was found across body sites.

This paper supports previous findings with regards to the commonality of generic factors. These factors help provide support to the feasibility of creating generic prognostic models such as the STarT MSK tool, a tool for stratification of patients with MSK pain (Campbell et al, 2016), and the Chronic Pain Risk Score, a chronic pain classification tool (Von Korff and Miglioretti, 2005) to predict patients at risk of poor outcome, and to the development of generic case-mix adjustment models for comparison of functional PROM data across patient cohorts and providers of MSK healthcare, with adjustment of PROM scores for those providers treating patients with a higher or lower risk of poor functional outcome compared to the average provider (Hart and Connelly, 2006, Coles, 2010).

A previous systematic review on case-mix adjustment models in MSK healthcare (Burgess et al, 2018, **Chapter 2**) identified two broad models; a UK National PROMS (NPROMS) model developed and validated in a UK secondary care orthopaedic setting (Coles, 2010, NHS England, 2012, NHS England 2013), and a US Focus on Therapeutic Outcomes (FOTO) model developed and validated in a community MSK setting (Resnik and Hart, 2003, Hart and Connolly, 2006). Variables highlighted within this review for use in case-mix adjustment modelling included; baseline functional status, age, gender, comorbidities, symptom duration, surgical history, payer, impairment type, index of multiple deprivation, ethnicity, assistance with questionnaire, and self-reported disability (Burgess et al, 2019, **Chapter 2**). This umbrella review provides further support to a number of variables already included in existing MSK case-mix adjustment models and has identified a number of additional independent predictors of functional outcome to consider for inclusion in future MSK case-mix adjustment modelling. These include symptom/pain severity, mental wellbeing, and BMI, which have strong evidence to support their inclusion as generic predictors, and consideration of inclusion of pain coping (including fear avoidance and

catastrophising), sick leave duration or compensation status (where relevant), vitality (measured using SF36/12), education/socioeconomic status (although already partly measured using IMD) and general health, which have moderate evidence to support their inclusion. The feasibility of collecting these predictors as part of a case-mix model within routine clinical data collection needs to be carefully considered, particularly for variables such as the mental health/wellbeing variable which was measured in a number of studies using the SF36 mental health component summary score, and vitality and general health which also used questionnaires (SF12, SF36). This may be considered too burdensome on patients to measure in some contexts and so shorter/briefer measures may be required.

The findings from this review have several implications. Existing case-mix adjustment models need testing to see if they can be modified using the predictors identified from this review to make them more applicable for a wide range of MSK conditions and clinical settings. Now that there are generic MSK outcome measures that have been validated for use across MSK conditions and settings such as the Arthritis Research UK Musculoskeletal-Health Questionnaire (MSK-HQ) (Hill et al, 2016), there is also the possibility of vastly reducing the number of routine clinical data variables collected through the use of generic outcome measurement and case-mix adjustment tools. This would reduce patient burden, reduce the complexity for clinicians in understanding and interpreting different measures, and be useful for commissioners/funders of research who typically pay for generic MSK services rather than specific MSK condition services and so want consistent data intelligence across the whole service they fund. Finally, having both a generic MSK case-mix adjustment model and outcome measure would enable a methodology to be developed to allow for fair inter-provider comparisons and benchmarking of MSK services, which at present is not available.

LIMITATIONS

The review was limited to English language studies due to the lack of translation services for non-English studies. It was also limited to search dates within the last 5 years although this has shown to be appropriate in this type of review (Aromataris et al, 2015).

The umbrella review examined all systematic reviews on predictive factors of functional outcome in MSK populations. There was therefore a risk of included studies including the same original cohort studies within their analyses. Overlap in reviews of systematic reviews is a recognised issue within umbrella reviews and decisions for inclusion can be; to include all identified studies and note the overlap, or, minimise overlap bias by specifying specific criteria and selecting the most comprehensive systematic review (Lunny et al, 2017). The latter approach however can lead to unintended loss of information through exclusion of important reviews (Lunny et al, 2017). Overlapping reviews within this umbrella review were therefore included and noted with regards to overall impact on results. For example, Struyf et al (2016) and Kooijman et al (2015), have three shoulder studies in common, with both finding strong evidence to support duration of symptoms and pain intensity as predictors of functional outcome in shoulder patients. We believe however that identified overlap will have had limited effect on the umbrella review results as we specified that evidence needed to be found in more than one study and more than one area of the body to be included as a generic predictor.

By including all systematic reviews including those evaluating specific treatments such as arthroplasty surgery alongside those looking more broadly at functional outcome in patients consulting for MSK conditions, there is a chance that some of the prognostic factors identified could actually be treatment effect modifiers (characteristics that influence the relationship between a specific intervention and outcome (Hill and Fritz, 2011)), these would need to be further evaluated in context to those interventions. This should have been largely avoided however due to factors needing to be predictive across body areas/conditions rather than for just one area and intervention type.

Conclusion

The umbrella review identified 21 systematic review articles meeting criteria for inclusion. All studies were of good quality. Following our high-level review of evidence, six generic predictors (baseline function/disability, symptom/pain severity, mental wellbeing, comorbidities, age and BMI) were found to have strong high quality consistent evidence across studies and anatomical body sites. Additional predictors (duration of symptoms, pain coping, workers compensation/sick leave, vitality, education, and general health) were also found to have moderate evidence across studies and body sites. All of these factors warrant consideration for inclusion within case-mix adjustment modelling of MSK outcomes. Next steps involve reviewing these findings alongside findings from the systematic review of existing MSK case-mix adjustment models (Burgess et al, 2019, **Chapter 2**), to determine a feasible (able to be captured simply at baseline) list of baseline patient factor variables for future testing in the development of a generic MSK case-mix adjustment model.

Review of candidate variables from the systematic review (Chapter 2) and umbrella review for use in case-mix adjustment.

VARIABLES FOR CONSIDERATION:

Case-mix adjustment modelling involves inclusion of all known patient factors that affect outcome within multivariate models in order for known confounders to be accounted for (Coles, 2010). Many models include widely accepted confounders, and then add those thought to impact on outcome or with less substantial supporting evidence. Variables are then tested independently for effect on outcome, and then if significant they are added into multivariable models using a stepwise approach, ensuring as they are added that any additional variables significantly improve model power (Coles, 2010). Key patient factor themes include those

described by Iezzoni (2009); patient demographics and characteristics, health related factors, patient attitudes and perceptions, and socioeconomic factors.

Fourteen studies were included within the systematic review of existing MSK case-mix adjustment models (see **Chapter 2** for full detail). **Table 2-5** demonstrates the most widely used variables across identified case-mix adjustment models (used in 3 or more included studies). The most significant variables within studies are also annotated, where the information was available. The most predictive and widely used variable across studies was baseline PROM score. Other important variables included comorbidities, surgical history, IMD, age, payer, duration of symptoms, impairment type, help with questionnaire, disability, gender, and ethnicity. All of these variables are feasible for widespread collection and warrant being considered for inclusion in MSK model development. Variables such as exercise history, living alone, FABQ, use of medication, and pain intensity had more limited support within the systematic review.

Twenty one studies were included within the additional umbrella review on independent predictors of MSK outcome. Following a high-level review of evidence, six generic predictors (baseline function/disability, symptom/pain severity, mental wellbeing, comorbidities, age and BMI) were found to have strong evidence across studies and anatomical body sites for predicting MSK functional outcome, supporting consideration for inclusion in case-mix adjustment. Additionally predictors of duration of symptoms, pain coping, workers compensation/sick leave, vitality, education and general health were also found to have moderate evidence across studies and body sites and therefore may also warrant consideration for inclusion. Considering the umbrella review alongside the case-mix adjustment model review we can see that there is as expected significant overlap in findings. For example, baseline function, comorbidities, age, duration of symptoms, and workers compensation/payer had agreement across reviews. IMD was a measure of socioeconomic status that was identified within the case-mix adjustment review and education within the umbrella review but ostensibly these could be considered to

measure the same construct. Symptom/Pain severity had strong evidence from the umbrella review supporting its inclusion, as did mental wellbeing and BMI. BMI has likely not been included previously in case-mix models due to difficulty with patients self-reporting weight and height accurately and/or inability to easily extract this from medical records. Mental wellbeing was frequently measured within prognostic studies using tools such as the SF12 and SF36 and therefore again could be considered too cumbersome for additional collection alongside other case-mix factors. Pain severity had only limited support from the case-mix adjustment review but when added to findings from the umbrella review it suggests that this factor should be included in any new case-mix model development.

There is a wide disparity in how variables are collated and entered into regression models, with a mixture of continuous, categorical and binary data across existing models. This will need to be explored and considered when looking to test and replicate case-mix models and when looking to analyse secondary data within the Research Institute.

VARIABLE DETAIL

Each variable warranting closer review (those variables used in 3 or more studies from the systematic review, and in 2 or more review studies across 2 or more pain sites within the umbrella review) is listed below with detail of how it was collected across included studies within the systematic review of existing case-mix adjustment models, followed by any additional evidence supporting inclusion or supporting additional variables from the umbrella review. Recommendations for those variables/patient factors that warrant initial inclusion in MSK case-mix adjustment model development based on current evidence are then detailed, alongside proposed detail for collection. This detail was considered with a view to informing the development of a future case-mix model to adjust community and primary care MSK outcome data, and secondly to inform a consensus process to gain clinician and patient feedback on

metrics and detail to include in a future large scale data collection (considering feasibility and acceptability for collecting in routine practice).

Baseline PROM score:

All systematic review studies included the baseline PROM score within the predictive model. This was categorised in different ways: Hart and Connolly (2006) coded the functional status (FS) into quartiles. Browne et al (2007) used the continuous variable of pre-operative score (Oxford Hip Score (OHS), Oxford Knee Score (OKS), EQ5D index score) as did other UK based studies (Coles, 2010, Gutacker et al, 2012, DoH, 2012, NHS England, 2013, Nuttall et al, 2015). Other models used a computerised adaptive tool (CAT) that gave a FS score estimate that transformed into a 0-100 (low to high functioning) continuous metric (Hart et al, 2011a, Hart et al, 2011b, Resnik et al, 2011, Yen et al, 2015, Gozalo et al, 2016, Werneke et al, 2016).

Function measured using a variety of self-reported PROMs was also found to have strong evidence to support its association with functional outcome in the umbrella review giving further support to the inclusion of this variable.

It can be seen that models included in the systematic review used different PROM tools to evaluate care outcomes, but most studies used a continuous variable of baseline PROM score within their case-mix adjustment model. Baseline PROM/functional score has strong evidence across studies to support its association with post-treatment outcome.

Recommendation: A continuous variable of baseline PROM score is strongly recommended for initial inclusion in MSK case-mix adjustment modelling.

Age:

All case-mix adjustment studies included age within the predictive model. This was categorised in different ways; Hart and Connolly (2006) used three categories of; younger (18-<45), middle age (45-<65) and older (65+), Gozalo et al (2016) categorized age into quartiles to allow for a non-

linear relationship. In contrast, UK studies used the patient's age in years (Browne, 2007, Coles, 2010, Gutacker et al, 2012, DoH, 2012, NHS England, 2013, Nuttall et al, 2015) as a continuous variable.

Age was found to have strong evidence to support its inclusion from the umbrella review providing further support to its inclusion within MSK case-mix modelling.

Recommendation: Age is strongly recommended for initial inclusion within MSK case-mix adjustment modelling. For model simplicity and following the UK NPROMS methodology above (NHS England, 2013), a continuous variable of age is supported.

Gender:

All systematic review studies included a binary item of gender in the predictive model, with UK studies highlighting gender as a key driver in a number of papers (Gutacker et al, 2012, NHS England, 2013, Nuttall et al, 2015).

Gender was found to have inconclusive/inconsistent evidence for predicting outcome within the umbrella review and the additional review did not support its inclusion within case-mix adjustment model development.

Recommendation: Gender has moderate evidence for inclusion in initial MSK case-mix adjustment model development on the basis of being found to be predictive within the case-mix adjustment review findings.

Comorbidities:

All systematic review studies included a measure of comorbidities in the predictive model, with the exception of Hart and Connolly (2006). Hart and Connolly (2006) considered comorbidities to be embedded in patient's perception of their ability to perform the FS items (baseline score) which they used as a measure of severity. Resnik et al (2011) assessed the addition of the Functional Comorbidity Index (FCI) to an established case-mix adjustment model. Using an

additive system was similar to previous studies in showing that discharge FS decreased as the number of comorbid conditions increased (Resnik et al, 2011). A weighted FCI or full list of comorbid conditions however, predicted slightly more variance than an additive approach (Resnik et al, 2011). Gozalo et al (2016), like Resnik et al (2011), used the Functional Comorbidity Index (FCI). Gozalo et al (2016) however categorised the FCI into low, medium low, medium high and high levels. Yen et al (2015) used a list based on the FCI that counted up to 30 comorbidities common to patients entering an outpatient rehabilitation clinic.

Browne et al (2007) used a list of 21 comorbidities but recommended the use of eight systemic conditions. Coles (2010) used patient-reported comorbidities (mixture of 12 systemic conditions), and health episode statistics (HES), and recorded comorbidities individually, and as part of the Charlson Index (includes 17 comorbid conditions). The latest iteration of the NPROMS methodology has moved to only including patient reported comorbidities with a list of 11 possible conditions (NHS England, 2013).

The umbrella review found strong evidence supporting the generic prognostic value of comorbidities and therefore strongly supports the inclusion of this variable within case-mix adjustment model development.

Recommendation: A measure of comorbidity is strongly recommended for initial inclusion within MSK case-mix adjustment modelling. For a UK population, use of the list of 11 conditions developed by the National PROMS team (NHS England, 2013) is recommended.

Duration of Symptoms:

Most studies included a measure of symptom duration. Coles (2010) used four categories for symptom duration (<1 year, 1-5 years, 6-10 years, >10 years), this variable was retained in the knee (OKS) model but not the hip (OHS) model or EQ5D models. The latest iteration of the national PROMs model uses symptoms for 2, 3 or 4 years in a yes/no binary list. Yen et al (2015) classed patients according to the number of days from onset to treatment (0-21, 22-90, >90).

Gozalo et al (2016) classed patients into acute (<21 days), subacute (22-90 days) or chronic (>90 days) as Yen et al (2015). The latest iteration of the US FOTO case-mix adjustment model expanded this to symptoms for; 0-7 days, 8-14 days, 15-21 days, 22-90 days, 91 days-6 months and over 6 months (Deutscher et al, 2018).

Duration of symptoms was found to have moderate evidence to support its association with outcome from the umbrella review therefore further supporting the inclusion of this variable in case-mix model development.

Recommendation: Duration of symptoms is recommended for initial inclusion within MSK case-mix adjustment modelling. Categories may include; 0-1 week, 1-2 weeks, 3-4 weeks, 4-5 weeks, 6-8 weeks, 9-11 weeks, 3-6 months, 6-9 months, 9-12 months, and greater than 12 months. This would form 10 initial categories that could then be condensed more in line with the US 6 categories if found to be as effective. This method is based on both the UK and US methods. The UK NHS population is likely to be more chronic in nature to the US private health service caseload. In the study by Gozalo et al (2016) for example, of 90,392 patients seeking outpatient therapy, 50.5% had had their symptoms/condition for more than 90 days duration and 36.6% for >183 days duration (6 months). In a recent UK study in a community MSK NHS setting, 48.4% of patients had symptoms for more than 6 months duration out of 484 patients (Comer et al, 2016), in an NHS primary care MSK triage service for more complex patients, mean duration of symptoms was 38 months (Sephton et al, 2010). Ensuring therefore that there is detail of those at the higher level of chronicity of 6-9, 9-12 and >12 months, is useful as well as having detail of those with more acute symptoms of 0-1, 1-2 and 3-4 weeks, to ensure a tool could work maximally across private and NHS settings and across acute and chronic MSK conditions.

Surgical History:

All systematic review studies included surgical history within the predictive model with the exception of NHS England's (2013) revised case-mix adjustment orthopaedic model. They

removed this variable due to separating out primary and revision procedures in case-mix adjustment models and performance analysis. Gutacker et al (2012) used number of coded secondary procedures from HES data and adjusted for revision surgery. Other UK NPROMS models used a binary variable of surgical history yes or no (Browne et al, 2007, DoH, 2012). Hart and Connolly (2006) coded surgical history as none, or, one or more as did Hart et al (2011a), Hart et al (2011b) and Werneke et al (2016). Yen et al (2015) used a categorical variable of, no surgical history, or, had a surgical history related to the impairment being treated. Gozalo et al (2016) categorised previous surgeries into none, one, or more. In the latest US FOTO case-mix adjustment model surgical history is a categorical list including; no related surgery, 1 related surgery, 2 related surgeries and 3 or more related surgeries (Deutscher et al, 2018).

Surgical history was not identified as a significant generic prognostic indicator from the umbrella review.

Recommendation: Surgical history is recommended for inclusion within MSK case-mix adjustment modelling based on case-mix adjustment review findings. For simplicity and due to the variety of methods across studies, an option of previous related surgery, yes or no, may be preferable.

Payer/Workers Compensation:

Within the systematic review, all US studies used insurance type or 'Payer' as an adjustment category. These included HMO, PPO, Medicare, Workers Compensation, Indemnity, Medicaid, and Other, insurance categories (Resnik et al, 2003, Hart and Connolly, 2006, Hart et al, 2011a, Resnik et al, 2011, Yen et al, 2015, Gozalo et al, 2016, Werneke et al, 2016)). Payer types have been used as proxy measures for a variety of demographic factors, including socioeconomic status, health status, access to healthcare resources, and healthcare providers attitudes toward patients (Burstin, 1992, Yen et al, 2015).

Within the umbrella review, workers compensation status/sick leave duration as a predictor of outcome was found to have moderate evidence.

Recommendation: For initial testing in a UK NHS based observational cohort, the payer or workers compensation variable would not be relevant as all patients would receive NHS funded care. For use across other healthcare systems, the payer variable would be strongly recommended for inclusion with categories dependent on the system itself such as the US FOTO model categories detailed above. Within a UK population sick leave duration could be used instead of workers compensation as it is supported by the umbrella review findings, or potentially benefit status could be used to try to mirror compensation status categories.

Impairment or Procedure Type:

Within the systematic review, Hart and Connolly (2006) used impairment type coded as the anatomical body part as did Gozalo et al (2016) who used a list of 10 body areas (lumbar, shoulder, knee, cervical, foot/ankle, hip, wrist/hand, elbow, ribs, craniofacial). The latest update to the FOTO model has taken out the impairment type variable in favour of separating out models into those for specific impairments, including a lumbar, elbow/wrist, hip, knee, neck, shoulder and general model (Deutscher et al, 2018). UK orthopaedic studies used procedure type to include primary and revision procedures (Coles, 2010, DoH, 2012, Gutacker, 2012, Nuttall et al, 2015) until revision procedures were separated out following methodology changes in 2013 (NHS England, 2013). NHS England (2013) included diagnosis codes in the updated model.

Impairment type/diagnosis was not found to be a generic prognostic factor from the umbrella review, but this may be due to the fact that most studies only looked at a single body/diagnostic area such as hip or knee OA.

Recommendation: For a generic case-mix adjustment model aiming to be used across MSK conditions, categories of body part/impairment are recommended. Categories may include; Head, Neck, Shoulder/upper arm, Lower arm/wrist, Hand(s), Upper back/chest/abdomen, Lower

back/pelvis, Hip/groin/thigh, Knee/lower leg, Ankle/foot, Other/not applicable, (10 categories) or a similar list of categories encompassing all major body areas or conditions.

Socioeconomic status:

All UK based systematic review studies included the Index of Multiple Deprivation (IMD) in case-mix adjustment modelling and found this variable to have a significant effect on predicting outcome. Browne and et al (2007) used the IMD based on inputted postcode using the 2004 Index. Coles (2010), Gutacker et al (2012), and DoH (2012), also used the income domain (IMD04i), within the 2004 Index. This was changed to the 2010 IMD Index within the NHS England (2013) update. No other forms of measuring socioeconomic status such as; education, occupation, health literacy, were utilised in case-mix studies but the 'Payer' variable above has also been linked to socioeconomic status (Burstin, 1992).

Within the umbrella review, education was found to have moderate evidence supporting its use as a generic prognostic factor and therefore its use within case-mix adjustment model development.

Recommendation: It is recommended that UK MSK case-mix adjustment models include a measure of socioeconomic status. The variable IMD as a measure of socioeconomic status based on the 2010 Index is supported by the UK National PROMs Programme (NHS England, 2013) in UK orthopaedic cohorts, or a measure of Education as supported by the umbrella review.

Exercise History:

Five US studies within the systematic review used an exercise history variable (Hart and Connolly (2006), Hart et al (2011b), Resnik et al, (2011), Yen et al (2015), and Werneke et al, 2016). Werneke et al (2016) found the exercise history variable was not significant in their observational cohort study and therefore this variable was removed from the model. Hart et al (2011b) also found limited effect from the exercise history variable. Hart and Connolly (2006) did find the exercise variable to be significant, but it was not one of the most significant variables, they

concluded that some variables included may be clinically insignificant. The latest FOTO model still includes exercise history as a categorical list of three including walking or jogging; at least 3x/week, 1-2x/week or, seldom or never (Deutscher et al, 2018). No UK studies used exercise history within their case-mix adjustment modelling.

Exercise history was not found to be a generic prognostic factor within the umbrella review and therefore does not provide support to its inclusion within case-mix adjustment model development.

Recommendation: The review evidence does not support the use of the exercise history variable in the case-mix adjustment of MSK PROM data.

Ethnicity:

Only UK studies in the systematic review used ethnicity as a case-mix adjustment variable (Coles, 2010, DoH, 2012, NHS England, 2013, Gutacker et al, 2012, Nuttall et al, 2015). Nuttall et al (2015) showed that patients recorded as Asian or Black on average have worse outcomes than those recorded as White. NHS England (2013) also found ethnicity to be a significant variable in primary hip and knee models.

Ethnicity was not found to be a generic prognostic factor for MSK outcome from the umbrella review.

Recommendation: There is moderate evidence to support the inclusion of ethnicity within initial case-mix adjustment model development. This variable is therefore recommended for inclusion where feasible. Categories are supported by the UK Office of National Statistics (ONS) and include: White, Mixed/Multiple ethnic groups, Asian/Asian British, Black/African/Caribbean/Black British, Other ethnic group (ONS, 2018).

Assistance with Questionnaire:

Within the systematic review, all of the NHS England NPROMS modelling includes assistance with questionnaire. In the latest iteration (NHS England, 2013), assistance with Q2 (post treatment questionnaire) was a key driver of outcome for primary hip and knee models. This reinforced previous findings on usefulness of this variable in Coles' (2010) inception paper and in the DoH (2012) paper. Nuttall et al also found assistance at Q2 to be a significant driver of outcome (Nuttall et al, 2015). None of the US FOTO papers included this variable and no further evidence was found to support this variable within the additional umbrella review.

Recommendation: There is moderate evidence to support the inclusion of the variable 'assistance with questionnaire' in MSK case-mix adjustment model development, it is therefore recommended for inclusion where feasible.

Disability:

Within the systematic review, all of the NHS England NPROMS modelling includes the variable of the patient classing themselves disabled at baseline. This is a binary question. NHS England (2013) found this variable to be significant in the primary hip model but not for the primary knee model. Nuttall et al found this variable to be a significant driver of outcome in a total knee arthroplasty cohort using the OKS.

Within the umbrella review self-reporting as disabled was not a factor evaluated within prognostic studies. Baseline PROM scores including those that measured disability were however found to strongly predict functional outcome.

Recommendation: It is recommended that a disability variable be initially included in MSK case-mix adjustment modelling, in view of findings from the National PROMs Programme (NHS England, 2013), and umbrella review. This could include the binary item of self-rated disability yes/no as included within the NPROMs method, or a disability questionnaire as used within umbrella review studies (Kooijman et al, 2015, Struyf et al, 2016).

Living Alone:

This variable was used across National PROMs Programme models (Coles, 2010, DoH, 2012, NHS England, 2013) within the systematic review, all papers found the variable to be significant in some models, but it was not a key driver of outcome in any of the models in any of the included papers. Living alone was not included as a variable in any of the US FOTO papers or within papers reviewed within the umbrella review.

This variable was not found to be a generic prognostic indicator of MSK functional outcome from the umbrella review.

Recommendation: Due to not being a significant driver of outcome in any papers, this variable is not recommended for inclusion in MSK case-mix adjustment modelling.

Fear Avoidance/Pain Coping:

Within the systematic review, Hart et al (2011a) developed a single item screening method used to classify patients by elevated versus not elevated fear avoidance beliefs of physical activity. They found that the addition of the single item screening method to an existing case-mix adjustment model predicting FS, improved the predictive ability of the model but only slightly (R^2 increased negligibly from 0.2997 to 0.3010). Gozalo et al (2016) classed fear avoidance simply as low or high based on FABQ-PA scores (high was classed >15 points). Hart et al (2011b) used a combination of psychosocial questionnaires including the FABQ-PA and the FABQ-WA to develop a variable for inclusion in an intake and change model. None of these papers however found the FABQ variable to be a key driver of outcome compared to other included variables. None of the included UK studies used the FABQ within their case-mix adjustment modelling.

Within the umbrella review 'pain coping' including fear avoidance was found to have moderate evidence supporting it being a generic predictor of MSK functional outcome.

Recommendation: This variable was found to have limited evidence supporting its inclusion in MSK case-mix adjustment model development and moderate evidence from the umbrella review and is therefore not recommended for initial inclusion.

Use of Medication:

Three US studies collected data on use of medication as a patient factor variable within the systematic review. Hart and Connolly, (2006), found this variable to have a significant impact on outcome but concluded that some variables although significant may be clinically unimportant. Hart et al, (2011b), and Werneke et al, (2016), did not find use of medication at intake for MSK problem (yes/no) to be significant in predicting outcome and therefore this variable was not retained in final case-mix adjustment models.

Use of medication was not found to be a generic prognostic factor of functional outcome within the umbrella review.

Recommendation: Due to the two more recent US FOTO papers not finding the use of medication to be a significant variable and no supporting evidence from the umbrella review of independent predictors, this variable is not recommended for inclusion in MSK case-mix adjustment modelling.

Pain Intensity/Symptom Severity

Pain was used for case-mix adjustment modelling in US papers by Hart et al (2011b) and Werneke et al (2016) within the systematic review. It was not found to be a significant driver of outcome and was removed from the baseline model in the study by Werneke et al (2016).

Within the umbrella review however, baseline pain/symptom severity/intensity was found to have strong evidence to support its ability to predict functional outcome, across studies and body sites.

Most studies used a Visual Analogue Scale (VAS) or Numerical Pain Rating Scale (NPRS) (0-10) (Struyf et al, 2016, Wilson et al, 2016, Werneke et al, 2016) to rate pain intensity.

Recommendation: Due to the findings from the umbrella review demonstrating strong evidence in support of pain intensity as a predictor of functional outcome, contrary to findings from the systematic review, this variable is recommended for inclusion in MSK case-mix adjustment modelling. A NPRS is supported to allow for a simple ordinal scale (0-10) (Werneke et al, 2016).

General Health

Self-reported general health was found to be a significant case-mix adjustment variable in studies by Browne et al (2007) and within the initial National PROMs case-mix model (Coles, 2010) but was not included within the later iteration when the model was reviewed and revised (NHS England, 2013). It was therefore not used frequently enough to be included in **Table 2-5** and initial recommendations from the case-mix adjustment systematic review. It is not clear from the NHS England (2013) paper why the general health variable was removed but it may have been due to collinearity as the comorbid conditions were updated. The general health variable was a categorical variable with 5 self-reported general health categories included (excellent, very good, good, fair and poor) (DoH, 2012).

General health was found to be a generic prognostic factor within the umbrella review and demonstrated moderate evidence in its ability to predict functional outcome across studies. The SF36 was however used to measure general health, and this could be argued to be too cumbersome for collection as part of a case-mix adjustment model to be used in day to day clinical practice.

Recommendation: The general health variable has limited data to support its consideration for inclusion in MSK case-mix adjustment model development and therefore is not recommended for initial inclusion.

BMI

No existing studies of case-mix adjustment of MSK PROM scores use body mass index (BMI) as an adjustment variable. This factor was however shown to strongly predict MSK functional outcome within the umbrella review. BMI can be self-reported by reporting height and weight although this may not be reported accurately by patients. Within orthopaedic services this information would often already be within the medical record allowing for data to be extracted and matched to self-report data, it is interesting however that this variable although strongly predictive within orthopaedic settings (3 of 4 studies supporting predictive ability of BMI set within orthopaedic surgical setting, and all including only clinical OA patients) is not included within the NPROMS case-mix modelling which may be due to difficulty with extracting from NHS HES records. This variable includes 4 categories (underweight (<18.5), healthy weight (18.5-24.9), overweight (25-29.9), obese (30-39.9)).

Recommendation: Evidence supporting this variable is limited to the umbrella review findings, but the umbrella review provided strong evidence of its predictive ability. This variable should therefore be recommended for inclusion based on umbrella review findings but will need discussion with experts within the field of MSK to discuss the feasibility of inclusion for everyday collection in clinical practice.

Mental Wellbeing

Depression was included within the 11 comorbidities used by the NPROMS case-mix model (DoH, 2012, NHS England, 2013). This comorbid condition was found to be a key driver of functional PROM outcome in both 2012 and 2013 iterations of the NPROMS case-mix model.

Within the umbrella review mental wellbeing was found to have strong evidence for predicting MSK functional outcome. Tools such as SF36 that were used to collect this detail however may be considered too burdensome for inclusion within a case-mix model.

Recommendation: Mental wellbeing/depression is recommended for inclusion within case-mix adjustment model development. As a minimum requirement, presence of depression should be collected as within the NPROMS model as a yes/no response (DoH, 2012, NHS England, 2013), with discussion with experts around practicality of using specific depression measures/PROMs.

Additional Patient Factor Variables:

Other variables utilised for case-mix adjustment within the systematic review studies were; referrer (Hart and Connolly, 2006), employment status (Resnik et al, 2003), and length of stay (Nuttall et al, 2015) but from the review findings of limited use of these variables they are not recommended for initial inclusion in case-mix model development. Full details of predictive variables explored within the umbrella review are shown in **Table 3-6** and those with evidence of generic predictive ability in **Table 3-7**. Vitality was shown to have moderate evidence of generic predictive ability from the umbrella review, this again was measured by the SF36 so may be too burdensome to include within a simple case-mix model, and widespread pain was found to have weak evidence of its ability to predict MSK outcome across conditions so is not recommended for inclusion.

VARIABLES SUMMARY

In summary, variables recommended for inclusion in MSK case-mix adjustment modelling with strength of recommendation are as follows:

Very strong evidence/Recommended (predictive across reviews (strong evidence from umbrella review)

Baseline PROM score (continuous score (PROM dependent)

Comorbidities (potential list of 11) to include depression/mental wellbeing

Age (continuous 0-120)

Strong Evidence/Recommended (predictive across reviews (moderate evidence from umbrella review)

Socioeconomic status/IMD (use of postcode for IMD 2010) or Education (categories not determined)

Surgical history (binary yes/no)

Payer type/Workers compensation (potential list of 10) or Sick leave duration (UK, categories not determined)

Duration of symptoms (potential list of 10)

Disability (binary self-rated yes/no or use of questionnaire)

Moderate Evidence/Recommended (predictive in either the case-mix adjustment review or umbrella review (strong evidence))

Impairment type (potential list of 11)

Pain intensity/severity (NPRS list of 11, 0-10)

BMI (4 categories)

Assistance with questionnaire (binary)

Gender (binary)

Ethnicity (potential list of 5)

Depression (binary) or Mental wellbeing (questionnaire)

Variables with limited support/Not recommended for initial inclusion in model development (used frequently in case-mix adjustment review studies (3 or more studies) but not significant predictor, or predictive in umbrella review (moderate evidence))

Living alone (binary)

Fear avoidance/Pain coping (FABQ)

Vitality (SF36)

Use of medication at intake (binary)

Exercise history (potential list of 3)

General health (potential list of 5)

Table 3-8 Recommendations for variables to include in MSK case-mix adjustment model development

Very strong evidence **Highly Recommended**	**Strong evidence** **Highly Recommended**	**Moderate evidence** **Recommended if feasible to collect in addition**	**Limited evidence** **Not Recommended**
Age (continuous, 0-120)	Disability (binary self-rated or questionnaire)	Assistance with questionnaire (binary yes/no)	Exercise history (potential list of 3)
Baseline PROM score (continuous)	Duration of symptoms (potential list of 10)	BMI (4 categories)	Fear avoidance (FABQ)
Comorbidities (potential list of 11)	Payer/Sick leave duration (potential list of 10 or sick leave duration list (to be determined))	Depression/Mental wellbeing (binary/mental wellbeing questionnaire)	General health (potential list of 5)
	Socioeconomic status (IMD 2010 or Education list)	Ethnicity (potential list of 5)	Living alone (binary yes/no)
	Surgical history (binary yes/no)	Gender (binary male/female (at birth))	Use of medication at intake for condition (binary yes/no)
		Impairment type/anatomical body part (potential list of 11)	Vitality (SF36)
		Pain intensity (NPRS 0-10)	

BMI; body mass index, FABQ; fear avoidance beliefs questionnaire, IMD; Index of Multiple Deprivation, NPRS; numeric pain rating scale, SF-36; short form 36.

Conclusion and Next Steps

The aims of this chapter were to; 1) carry out an umbrella review of systematic reviews on predictors of MSK functional outcome, 2) explore the findings from **Chapter 2** in more depth with regards to potential case-mix adjustment model variables alongside any additional predictors identified in the umbrella review, to derive strength of evidence supporting inclusion of identified variables within future MSK case-mix adjustment modelling. Both of these objectives have been completed with the required output of generating a list of generic candidate variables

recommended for inclusion within future MSK case-mix adjustment modelling and providing potential detail for future data capture (see **Table 3-8**).

A limitation of both reviews (systematic and umbrella) was the number of different PROM measures used across studies. Each measure requires a validated case-mix adjustment model, as different measures of functional outcome may require different variable sets, as demonstrated within the NPROMS methodology (Coles, 2010). The use of a generic MSK PROM measure such as the newly developed MSK-HQ (Hill et al 2016) across MSK services and pathways, could allow for the development of a single case-mix adjustment tool that could be tested and validated across MSK conditions and settings allowing for a more feasible and realistic model for regular and widespread clinical application.

Next steps:

To use the recommendations for case-mix adjustment variables to inform consensus work developing consensus around a core MSK outcome set for use in routine practice with a view to making recommendations based on best evidence for a feasible case-mix adjustment model for use across MSK services alongside optimal PROMs and PREMs **(Chapter 5)**.

To test identified variables and models, with regards to variable independent predictive ability using univariate regression models, and collective predictive ability using multivariate regression models in a community/primary care MSK dataset (TAPS/STarT MSK Main Trial) **(Chapter 6)**.

4 Chapter 4: Costing community and primary care musculoskeletal services.

The aims of this chapter were to; a) review costing methodology in the context of MSK services and community/primary care, and; b) undertake a systematic review to identify key cost indicators/drivers in MSK economic analyses undertaken within these settings.

Contribution to the Thesis

- Accurate and comparable cost data is fundamental in order to improve the value of NHS MSK services.
- NHS costing systems are complex but are moving towards more accurate patient level costing methods.
- Key resource use variables for MSK healthcare have been identified and include; GP visits, outpatient medical specialist visits, and physiotherapy visits.

This Chapter has been published in part in:

Burgess, R., Hall, J., Bishop, A., Lewis, M. and Hill, J., 2020. Costing Methodology and Key Drivers of Health Care Costs Within Economic Analyses in Musculoskeletal Community and Primary Care Services: A Systematic Review of the Literature. *Journal of primary care & community health*, *11*, p.2150132719899763.

Review of MSK costing in UK community and primary care

Musculoskeletal (MSK) conditions account for the third largest area of NHS programme spending (£4.7 billion in 2013/14 (NHS England (2014a), ARUK, 2018)). Following Lord Carter's review for the Department of Health in 2016 (Carter, 2016), there has been an increased focus on high-cost procedures and interventions, and on defining, mapping, and collecting cost data. This focus includes areas such as elective surgical procedures and MSK trauma admissions (Carter, 2016, NHS Improvement, 2016a, NHS England, 2021a). There has been significantly less attention given

to defining costs and collecting widespread comparable MSK cost data in primary care and community settings. There is however widespread agreement on the need to re-design MSK care pathways such as access to orthopaedic elective care to make better use of community and primary care services such as clinical triage (NHS England, 2017), and implement new approaches such as the First Contact Practitioner (FCP) model (NHS England and NHS Improvement, 2019). Accurate community and primary care cost data is necessary as system level changes are adopted and MSK healthcare traditionally provided in secondary care settings is instead delivered in community and primary care clinics and work previously carried out by GPs is instead undertaken by Advanced Physiotherapists working in FCP roles. With limited NHS resources, any shift in core activity needs to be appropriately costed and shown to be both clinically effective, cost effective, and sustainable, to fully inform transformation and cost improvement plans.

Unwarranted variation within the NHS acute sector has been estimated at a value of £5bn in terms of an efficiency opportunity (Carter, 2016). This was estimated by reviewing all key resource areas such as; clinical staff, pharmacy and medicines, diagnostics and imaging, procurement, back office functions and estates and facilities (Carter, 2016). Identifying variation requires the use of these specific metrics as outlined in Lord Carter's report. There has been a large focus on costing and efficiency in acute care, leading to clear examples of unwarranted variation being identified within acute services. Hip prostheses for example, have been shown to vary in cost from £788 to £1590 across NHS Trusts, with those buying the most not paying the least (Carter, 2016). This work is developing within the Model Hospital/Model Health System (NHS Improvement, 2016a, NHS England, 2021a), a digital information service set up initially to support secondary care providers to improve productivity and efficiency. The Model Hospital demonstrates how appropriately applied metrics can allow for a useful comparison of services and their efficiency. However, similar economic metrics have not yet been formalised or evaluated for MSK community and primary care in the UK.

This Chapter focused on identifying important resource use metrics/indicators for MSK community and primary care services, where evidence is currently lacking. Identified indicators were then put forwards to stakeholder consensus (**Chapter 5**) to agree a feasible list that could be used to identify cost variations in key resource areas within MSK community and primary care with consideration to capture in routine practice.

NHS COSTING

Understanding the cost of providing NHS care is vitally important both locally and nationally, in order to make necessary decisions about how to deliver high quality sustainable services for the future across the UK (Healthcare Financial Management Association (HFMA), 2016). With an increased prevalence of long-term conditions and an ageing population, the NHS needs to develop new models of care to meet the increasingly complex health needs of the population (NHS England, 2014b). With increased NHS funding set out in the NHS Mandate (2018), comes increased responsibility for the NHS to minimise waste and make best use of available resources, with the mandate setting out a clear responsibility to reduce unwarranted variation in healthcare delivered. Good cost data is therefore essential in helping NHS organisations to understand variation in treatment and resources in order to plan future sustainable models of care (HFMA, 2016).

The HFMA (2016) define costing as; *'the quantification, in financial terms, of the value of resources consumed in carrying out a particular activity or producing a certain unit of output.'* Costing involves being clear about activities to be costed in terms of defining what cost is being sort and ensuring that everything and everyone involved in carrying out that activity are included within the costing calculation (HFMA, 2016).

Costing information is used in a variety of ways. One of which is to support value-based decision making (HFMA, 2016). Achieving high value for patients is integral to a successful health system,

with value defined as the health outcome achieved per monetary unit spent (Porter, 2010). Shifting the focus from volume of services delivered, to value of services provided involves transparent measurement of outcomes and costs and is integral in allowing NHS organisations to transform in order to run more effectively and efficiently (Porter, 2010, HFMA, 2016). Costing information also helps managers to monitor and manage budgets appropriately, develop business plans that are realistic and fit for purpose, and can also be used to compare and benchmark services, highlighting potential investment opportunities (HFMA, 2016, NHS RightCare, 2016).

Accurate and consistent collection of cost data is therefore important to allow for meaningful use of costing information (HFMA, 2016). NHS Improvement issues 'Approved Costing Guidance' annually, giving recommendations and guidance with regards to regular cost collection and reporting (NHS England and NHS Improvement, 2020, NHS England, 2021b). Before reviewing costing guidance and practice in detail, it is first useful to consider payment practices across the NHS to consider the wider context of NHS finance.

PAYMENT SYSTEMS

Since the NHS was established block contracts have been the predominant payment system, with this continuing to be the case across Scotland, Wales and Northern Ireland (British Medical Association (BMA), 2021). Block contracts are payments made in advance to deliver a specific service. Payments are made regularly for a set time period to deliver an estimated amount of activity. Advantages are that payments are timely and predictable, but disadvantages include payments not being flexible to meet the demands of changes to patient numbers or provider costs. They also do not incentivise improved care or efficiency (BMA, 2021).

Payments made to the acute sector are currently dominated by the National Tariff, otherwise known as 'payment by results' (PbR) (BMA, 2021). Diagnosis Related Groups (DRGs) are used to

classify and determine pricing, and in the UK are known as Health Related Groups (HRGs). HRGs form the basis for the National Tariff, and there are currently 1400 mandatory tariffs. The National Tariff was introduced in 2003/2004 to allow for choice and competition within the NHS (BMA, 2021). The advantages to the National Tariff are that it links payment directly to the activity and services actually delivered, whilst recognising complexity and severity of treatment provided (HFMA, 2016). A disadvantage is that it was setup to fundamentally change the way secondary care services are funded, rather than focussing on the integrated health system with regards to the inclusion of community and primary care services (HFMA, 2016).

The use of the National Tariff may start to diminish with changes to care models set out in the Five Year Forward View (FYFV) (NHS England, 2014b) and the NHS Mandate (2018), to develop 'multi-speciality community providers' (MCPs) and 'primary and acute care systems' (PACS), to form more integrated provider models. Most sustainability and transformation plans in England now aim to move towards an outcome based capitated budget approach (BMA, 2021). Capitation involves a 'lump sum' payment being made to care providers based on the number of patients in a target population. This approach is used to determine core funding in UK general practice but is uncommon presently outside of primary care (BMA, 2021). The main advantage to the capitated approach is the ability to facilitate more integrated care, which has been shown to help professionals work more closely together due to one budget funding all of the patient's care (BMA, 2021, Struijs and Baan, 2011).

REFERENCE COSTS

Reference costs provide the average unit cost to the NHS of providing defined services within a given financial year. They are the richest source of financial data available about the NHS, providing detailed comparisons of costs across NHS organisations (HFMA, 2016). Data collection is nationally mandated. Reference Costs 2016/17 (NHS Improvement, 2017) details how 234 NHS providers in England spent £66.2bn delivering healthcare in 2016/17. Quality of this powerful

data resource is however dependent on individual trusts internal costing processes. An audit by NHS Improvement in 2015 (Monitor, 2015) showed that 49% of acute trusts audited had made materially inaccurate reference cost submissions, demonstrating the complexity of producing accurate costing data (HFMA, 2016). Reference costs were designed for central purposes, with the primary objective of allowing for the calculation of a national tariff (Chapman and Kern, 2010). Calculation of reference costs requires a top-down cost-calculation process including allocating overheads across a number of HRGs. Calculations are complicated and unclearly linked to clinically relevant categories (Chapman and Kern, 2010). They do not provide detailed information on how to understand organisational costs, and therefore the output is not something easily understood by clinicians making it difficult for front line staff to use the cost information to drive performance and value for money (Chapman and Kern, 2010).

PATIENT LEVEL COSTING (PLICS)

Accurate and comparable cost data is fundamentally important if providers are to identify and improve outcomes for patients, developing sustainable services for the future (NHS Improvement, 2016b, NHS England and NHS Improvement, 2020). The aim set out within the FYFV (NHS England, 2014b) is to deliver new ways of working (MCPs and PACS) and improve costing processes throughout the NHS (NHS Improvement, 2016b). To make cost information more credible and help organisations understand their NHS business model better and to better inform national collection of cost data and PbR tariff setting, NHS Improvement (2016b) has developed a new costing matrix called the; Patient-Level Information and Costing System (PLICS) (Chapman and Kerr, 2010).The process of producing and collecting the newly developed PLICS data is set out within the Approved Costing Guidance documents (NHS England and NHS Improvement, 2020, NHS England, 2021b), and will become mandated across all providers by 2020/21 (HFMA, 2016, NHS Improvement, 2016b, NHS England and NHS Improvement, 2020). Sector specific digital platforms are being launched to help early implementers, with the aim of

reducing the burden of collection of cost data (NHS Improvement, 2016b). Costing education and training is highlighted as a priority area that will also become integrated into the PLICS approach. Reference costs are currently still mandated, whilst the PLICS approach is being developed. The PLICS process aims to provide a single costing collection process across acute trusts, mental health and ambulance trusts, and community service providers (NHS Improvement, 2016b, HFMA, 2016). PLICS data will; support the calculation of national tariff replacing reference costs, allow for analysis of cost data against peers, and be collectively analysed by teams within NHS Improvement (NHS Improvement, 2016b). The costing transformation programme will be overseen by NHS Improvement, and a costing assurance programme (CAP) will be put in place to ensure accuracy of submitted data and demonstrate the successful implementation of the costing process changes (NHS Improvement, 2016b, NHS England and NHS Improvement 2020).

NHS Improvement (NHS England and NHS Improvement, 2020) set out four key objectives of collecting information for costing using PLICS, which include; ensuring providers collect the same information for costing and collection purposes and comparison with peers, to ensure allocation of the correct quantum of cost to the correct activity, to ensure accurate matching of costed activity to the correct patient contact, and to support local reporting of cost information by activity by providers within business intelligence dashboards. Classification of costs includes: patient-facing costs which are those that relate directly to the delivery of patient care and are driven by patient activity, these will be both pay and non-pay, and support costs which do not directly relate to delivering patient care but to running the organisation (eg board costs, HR, finance, estates), or service level support costs such as ward clerks and service management costs (NHS England and NHS Improvement, 2020, NHS England, 2021b).

THE MODEL HOSPITAL

Lord Carter (2016) endorsed the creation of a Model Hospital to allow for development of underlying metrics and benchmarking practice. The aim was for clearer methodology to be

applied across providers in order to maximise productivity and effectiveness whilst minimising costs (NHS Improvement, 2016a, NHS England, 2021a). The process of optimising resource use involves the creation of preferred metrics for comparison. These metrics are being developed iteratively within the Model Hospital to set out productivity and efficiency. Costing metrics include; adjusted treatment cost (ATC) used to see how trusts vary in cost for a given output, and weighted activity units (WAU) used to compare productivity across providers (NHS Improvement, 2016a). The portal is still in early development and is continually evolving to meet the requirements of providing a nationally accessible performance information system. It aims to help reduce unwarranted variation by highlighting performance metrics including productivity, efficiency and quality of care. Initial focus has been on acute sector metrics with community services data still under development within the 'Model Health System' (NHS England, 2021b).

UNIT COSTS OF HEALTH AND SOCIAL CARE

The unit costs for health and social care publication is updated annually, and provides estimates of service and staff costs across health and social care including community and primary care costs (Curtis and Burns, 2017). Unit costs for community-based scientific and professional staff for instance are included with a breakdown of each Agenda for Change band including basic pay and additional costs such as qualifications and capital overheads (Curtis and Burns, 2017). Unit costs are also provided, such as an hour of GP or practice nurse time (Curtis and Burns, 2017). This approach to unit costing is grounded in economic theory and underpins costing estimates within research based economic analyses by NICE in assessing costs of interventions, and by the Department of Health in informing policy (Curtis and Burns, 2017).

NHS RIGHTCARE

The primary objective of NHS RightCare is to maximise the value that patients derive from their own care and treatment and the value that the population derives from investment in healthcare

(NHS RightCare, 2016). Commissioning for value focus packs for MSK conditions, trauma and injuries provide a top-down view of NHS spend on MSK care and link spend to outcome in the areas of elective care, fractures and falls. This information on health gain achieved for pound spent is not however available for the treatment of back, neck and MSK pain in community and primary care. This makes it difficult to fully ascertain the value of the entire pathway of MSK care, as costs are not provided at a patient level and therefore do not allow for detailed analysis of the key drivers of variation in costs across CCGs. NHS RightCare focus packs would benefit from increased data metrics to support the process of evidencing and maximising value of MSK services across the whole MSK pathway.

SUMMARY

Although new NHS costing frameworks aim to reduce the burden of collecting cost data whilst improving consistency and accuracy of reporting performance, it is clear that NHS costing systems remain complex. Whilst the PLICS system at a patient level aims to improve standardisation and breadth of costing metrics collected, it will take time for this process to be fully integrated into NHS systems and fully evaluated to ensure it has been successfully embedded and can deliver on the objectives set out within the Approved Costing Guidance documents (NHS Improvement, 2016b, NHS England and NHS Improvement, 2020). It is therefore useful in the short-term to identify if a more limited number of resource variables could be used within the context of MSK community and primary care in order to compare resource utilisation whilst minimising collection and reporting burden, allowing for data to be easily interpreted by all MSK stakeholders for use in developing efficient models of MSK care.

Previous Literature review on costing methodology

In 2005 Mogyorosy and Smith (2005) undertook a literature review exploring the main methodological issues in costing healthcare services by examining the scientific literature on methodologies for calculating health service costs. The review set out to examine methods used

to estimate costs associated with delivering healthcare services at the micro-level, in both in-patient and out-patient settings, providing a detailed summary and international comparison. This review was part of the wider HealthBASKET Project detailing health benefits and service costs in Europe (European Health Management Association (EHMA), 2007). Key findings from the EHMA (2007) literature review are summarised below:

There was no universally accepted costing methodology between the nine member states. Several methods were identified to estimate the unit costs of services. Basic principles of costing were agreed as defining the; objectives of costing, perspective, time horizon, and description of the service to be costed. Methods for costing included; the identification of resources used to deliver a service, the measurement of resource utilisation in natural units, and attaching monetary value to the resource used. Costing services is a trade-off between accuracy of cost information and costs and practicality of obtaining the information. There are five broad ways to value resources; direct measurement of costs, cost accounting methods, standard unit costs, fees/charges/market prices, and estimates/extrapolations.

Pertinent to the aims of this review of the literature around costing, Mogyorosy and Smith (2005), also highlighted the usefulness of defining cost drivers. Cost drivers are variables, such as medical visits, that affect costs over a given period of time and can be directly linked to changes in costs (Horngren, 2003). It is important to identify the most relevant cost drivers as these can help explain changes or differences in overall costs of care. It is also sometimes sufficient and necessary to limit data collection (or costing) to the key cost-drivers (Johnston, 2001). Activity based costing looks to improve the accuracy of cost estimation by using multiple cost drivers (Mogyorosy and Smith, 2005). This supports the notion of identifying key cost drivers in MSK community and primary care services in order to allow for practical collection and appropriate comparison of MSK community and primary care healthcare costs, that could include elements of the NHS PLICS costing requirement but concentrate on the specific drivers in community and

primary MSK care that drive the majority of service level costs. The next step is therefore to analyse specific MSK economic analyses in order to identify the pertinent cost drivers that drive costs in MSK community and primary care. A systematic review of MSK health economic analyses based primarily in community and primary care will therefore be carried out in order to identify costing methods and key cost drivers within the economic literature.

Costing methodology and key drivers of healthcare costs within economic analyses in MSK community and primary care services: A systematic review of the literature.

Abstract

Background: Identifying variation in MSK service costs requires the use of specific standardised metrics. There has been a large focus on costing, efficiency, and standardised metrics within the acute MSK setting, but far less attention in primary care and community settings.

Objectives: a) to assess the quality of costing methods used within MSK economic analyses based primarily in primary and community care settings, and, b) to identify which cost variables are the key drivers of MSK healthcare costs within these settings.

Methods: Medline, AMED, EMBASE, CINAHL, HMIC, BNI, and HBE electronic databases were searched for eligible studies. Two reviewers independently extracted data and assessed quality of costing methods using an established checklist.

Results: 22 studies met the review inclusion criteria. The majority of studies demonstrated moderate to high quality costing methods. Costing issues included studies failing to fully justify the economic perspective, and not distinguishing between short and long run costs. Highest unit costs were; hospital admissions, outpatient visits and imaging. Highest mean utilisation was; GP visits, outpatient visits, and physiotherapy visits. Highest mean costs per patient were; GP visits, outpatient visits, and physiotherapy visits.

Conclusion: This review identified a number of key resource use variables that are driving MSK healthcare costs in the community/primary care setting. High utilisation of these resources (rather than high unit cost) appears to be the predominant factor increasing mean healthcare costs. There is however need for greater detail with capturing these key cost drivers, to further improve the accuracy of costing information.

Introduction

Economic evaluation is 'the comparative analysis of alternative courses of action in terms of both their costs and consequences' (Drummond et al, 2015). The objectives of health economic analyses are to identify, measure, value, and compare costs and consequences of alternative treatments/pathways of care (Drummond et al, 2015). Economic evaluation techniques therefore provide the framework for capturing costs and outcomes/benefits of different health interventions (Charles and Edwards, 2016). In this review economic costing methodology was explored, which aimed to identify the most important cost variables for making resource use comparisons within primary and community MSK settings. The quality of costing methods was also evaluated to determine the strength of these findings.

Graves et al (2002) described 12 criteria for specifically assessing the quality of costing methods of economic analyses. These criteria include for example; stating and justifying the perspective, distinguishing between short and long run costs, and reporting methods for estimating quantities of resources. Graves et al (2002) reviewed 45 economic analyses against the 12 criteria in 2002 and concluded that more attention should be given to costing methods to ensure the accuracy of costing estimates.

Mogyorosy and Smith (2005) also reviewed methodological issues with costing healthcare services across Europe in 2005 and highlighted the usefulness of defining cost drivers. It is important to identify the most relevant cost drivers for a specific setting as these can help explain changes or differences in overall costs of care.

Costing within economic analyses will differ dependent on the perspective, which can include a health perspective looking at direct costs to health managers, or a societal perspective looking at direct and indirect costs, including productivity losses (Jackson, 2012). In this review only direct healthcare costs were examined in detail as the review looked to identify key drivers of MSK healthcare costs only.

The review aimed therefore to evaluate the quality of costing methods used within MSK economic analyses, and also to identify the most important cost drivers (provider variables) that could be reported by individual primary care and community MSK services and used to develop a system level methodology to broadly cost and compare these services. The perspective was that of a health service perspective looking at costs incurred by MSK patients being predominantly treated in Community/Primary Care, as this is the area at present where a clear methodology is lacking.

Method

The review followed protocol guidance set out within the PRISMA statement for systematic reviews (Moher et al, 2015).

ELIGIBILITY CRITERIA

Inclusion criteria were: all types of economic analyses, economic analysis needed to be the primary focus of the study and clearly stated from the outset, studies published within the last 10 years (January 2008 to May 2018) in order to ensure applicability of costing approaches, English language papers only (due to resource limitations of the review), studies primarily set within community or primary care health settings to support the focus of the paper, patient population to include patients undergoing treatment for most prevalent MSK disorders (back, neck, shoulder, knee or all MSK conditions (Urwin et al, 1998, Jordon et al, 2010), studies needed to

report disaggregated healthcare costs, and papers needed to be published in full within a peer reviewed journal.

SEARCHES

Medline, AMED, EMBASE, CINAHL, HMIC, BNI, and HBE electronic databases were searched. Searches were limited to papers within the last 10 years only to ensure applicability when looking to review costing methods. Search terms were grouped into core areas of; type of economic evaluation such as cost benefit or cost utility (Charles and Edwards, 2016), study type such as analysis or evaluation, and prevalent musculoskeletal condition such as back, neck, knee, or shoulder (Urwin et al, 1998, Jordon et al, 2010). Searches were combined using Boolean logic (AND and OR). Additional grey literature was identified using reference lists of included articles. Included studies were required to have a combination of key search terms (**Table 4-1**) within the study title, to ensure studies were focussed on economic evaluation, and not on intervention effectiveness.

Table 4-1 Search term

Column 1 (economic evaluation type)	Column 2 (study type required)	Column 3 (MSK conditions)
Economic Cost consequence Cost benefit Cost utility Cost effectiveness Cost minimisation/minimization	Analysis Evaluation	Musculoskeletal Back Neck Spinal Knee Shoulder

The review did not set out to identify every economic analysis within the field of MSK, but to systematically identify economic analyses within the field of MSK from community/primary care settings, that met the inclusion criteria, using a structured, transparent, and reproducible search methodology. The full electronic search strategy for searching databases can be viewed within the appendix (see **Appendix 4-1**).

SELECTION CRITERIA

One reviewer (RB) reviewed search results and identified those based on title and abstract unquestionably excluded from the review, and those appropriate for full review. The same reviewer (RB) and a second reviewer (JH) then independently read full papers for identified studies and determined eligibility based on inclusion/exclusion criteria to ensure all criteria were met for inclusion into the systematic review.

DATA EXTRACTION

Data was extracted using an extraction sheet made up of study descriptors such as author, year, title, objectives, setting, population, study size, perspective, outcomes, cost variables, cost year, costing sources, and analysis type e.g. CEA, CUA, CBA, CCA (cost effectiveness/utility/benefit/consequence analysis respectively). This was followed by the 12 quality criteria described by Graves et al (2002). Two reviewers independently extracted data (RB and JH), one reviewer had a clinical MSK background (RB) and one was a health economist (JH). All healthcare costs were extracted and included whether the cost was for treatment within primary, community or secondary care as although the studies were predominantly set within primary care/community clinics, costs were still borne by the healthcare system across all settings (for example if a patient was sent for an investigation or consultant opinion).

The focus of the study was restricted to costs and therefore outcomes and specific interventions were not reviewed. Detailed costs were reported and analysed for the usual care/control groups only in order to allow for broad comparisons of key cost drivers to be made across studies. Due to the broad focus of the review (including all common MSK conditions) cost drivers were ranked for each study rather than being directly compared.

QUALITY OF COSTING METHODOLOGY

Quality of costing methods within included studies was evaluated using 12 criteria described by Graves et al (2002) and following this guidance organised into 4 pre-defined categories; general costing issues (criteria 1-4), methods used to determine the quantities of resources (criteria 5-7), methods used to determine the value of resources consumed (criteria 8-9), and reporting of data (criteria 10-12) (Graves, 2002). Quality of costing methodology within studies was evaluated independently by two reviewers (RB and JH) and inputted directly into the extraction sheet. A third reviewer was then available (ML) if agreement could not be reached.

Results

SEARCH RESULTS

751 records were identified through electronic searching of relevant databases. This gave 344 records after electronic duplicates were removed (see **Figure 4-1** for detail).

Following initial review of titles and abstracts 276 articles were excluded for failing to meet the eligibility criteria. Following full review of the remaining 68 articles against eligibility criteria by the two reviewers, 22 studies were suitable for inclusion in the review. The main reasons for exclusion were; not providing disaggregated costs to determine patient-level healthcare costs, not being undertaken from the right clinical setting, being a secondary analysis rather than a pre-planned economic analysis, being an abstract only, or being a model or review rather than a primary study.

The majority of studies were economic analyses conducted alongside RCTs (n=21) or other trial-based analyses (n=1) (Whitehurst et al, 2015). See **Table 4-2** for characteristics of included studies.

Figure 4-1 Flowchart of search results

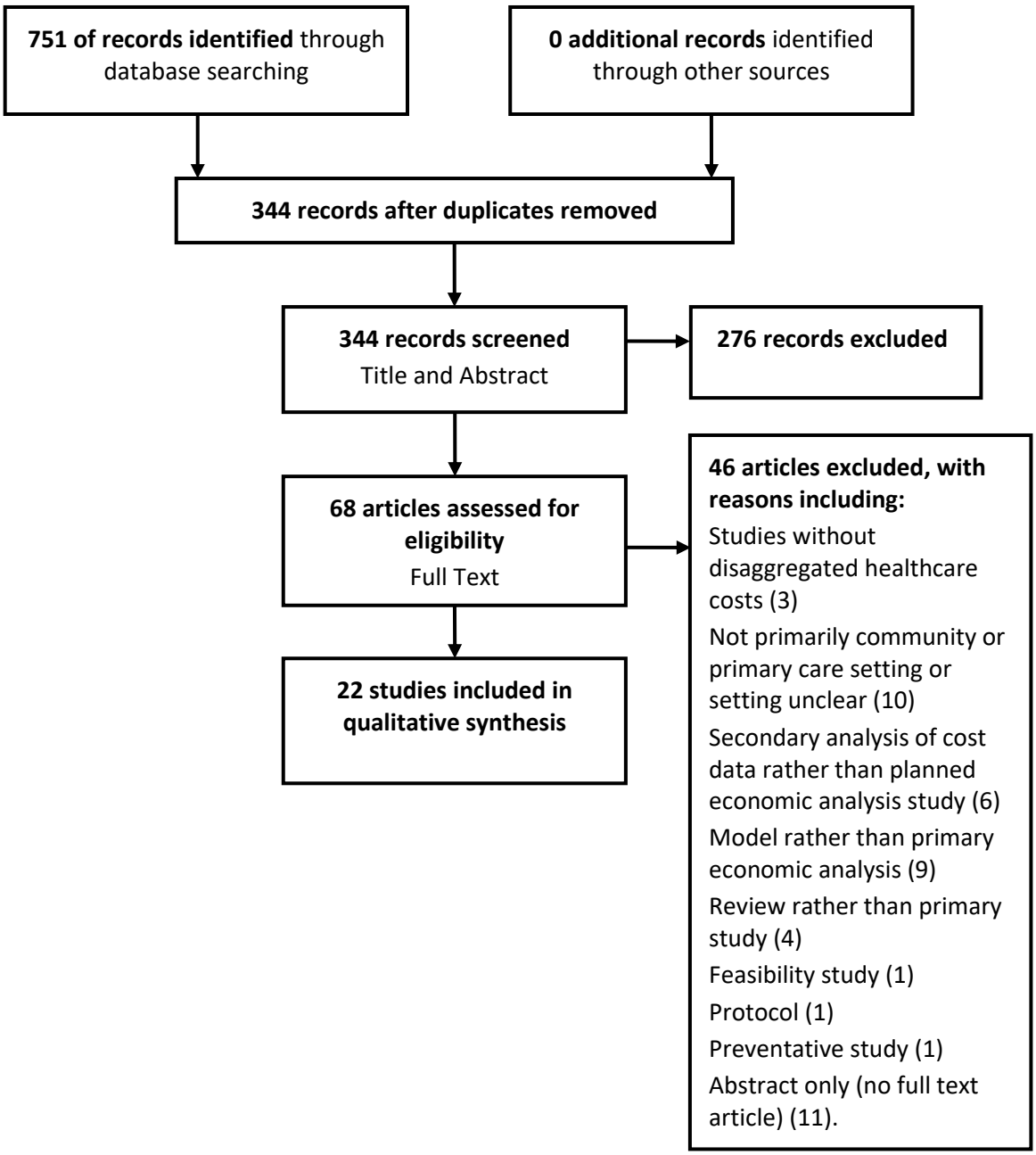


Table 4-2 Characteristics of included studies

Author (first)	Country/Origin	Population	RCT	Setting	Study Size (patients)	Analysis type	Group costs reported for in analysis
Barton 2009	UK	OA knee with high BMI (>28)	Yes	PC	389	CUA	Usual care
Bosmans 2011	Netherlands	Subacute neck pain	Yes	PC and C	146	CUA	Usual care
Bultmann 2009	Denmark	MSK disorder	Yes	C	119	CBA	Usual care
Chuang 2012	UK	Chronic LBP	Yes	PC and C	313	CEA	Usual care
Essex 2017	UK	Neck pain	Yes	PC and C	293	CEA, CUA	Usual care
Haines 2017	Australia	Chronic LBP	Yes	PC and C	112	CEA	Usual care
Henchoz 2010	Switzerland	Chronic LBP	Yes	PC and C	105	CUA	Usual care
Hollinghurst 2008	UK	Persistent back pain	Yes	PC and C	579	CUA	Usual care
Hollinghurst 2013	UK	MSK disorder	Yes	C	2249	CCA & CUA	Usual care
Jenson 2017	Denmark	LBP	Yes	PC	1101	CUA	Usual care
Jowett 2013	UK	Shoulder	Yes	PC and C	232	CUA	Usual care
Lamb 2010	UK	Subacute or chronic LBP	Yes	PC	701	CUA	Usual care
Lambeek 2010	Netherlands	Chronic LBP	Yes	PC and SC	134	CUA	Usual care
Marra 2014	Canada	Knee OA	Yes	PC	139	CUA	Usual care
McKenna 2009	UK	Acute shoulder Pain	Yes	PC	200	CEA, CUA	Usual care
Pinto 2013	New Zealand	Knee and hip OA	Yes	PC and C	206	CEA & CUA	Usual care
Saha 2018	Swedan	Neck or back pain	Yes	PC and C	352	CEA, CUA	Usual care
Smeets 2009	Netherlands	Chronic LBP	Yes	C	172	CEA, CUA	Combined treatment
van de Roer 2008	Netherlands	Chronic LBP	Yes	C	114	CEA & CUA	Usual care

Vermeulen 2013	Netherlands	MSK disorder	Yes	C	163	CEA	Usual care
Werner 2016	Norway	LBP	Yes	PC	216	CEA	Usual care
Whitehurst 2015	UK	LBP	No	PC	922	CUA	Usual care

C; community, CEA; cost effectiveness analysis, CBA; cost benefit analysis, CUA; cost utility analysis, LBP; low back pain, MSK; musculoskeletal, OA; osteoarthritis, PC; primary care, SC secondary care

Table 4-3 Quality of costing methodology in included studies (criteria defined by Graves et al, 2006).

	Graves Costing Quality Criteria													
	1	2	3	4	5	6	7	8	9	10	11	12	Total	Complete
Author	Perspective stated	Perspective justified	Cost data satisfied perspective	Short and long run costs (as appropriate)	Methods for quantities resources used by patients	Methods for allocating time of human resources	Methods for allocating other resources (fixed) between patients	Methods for estimating price, unit costs, charges	Were data other than hospital charges used	Was the year data collected reported	Was base cost year reported	Were adjustments made for different time periods where necessary	Total score out of 12	Complete cost data (%)
Barton 2009	Yes	No	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	10	48%
Bosmans 2011	Yes	Partial	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	10.5	81-86%
Bultmann 2009	Yes	Partial	Yes	No	Yes	Unclear	Unclear	Partial	Yes	Yes	Yes	Yes	8	55-82%
Chuang 2012	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	11	unclear
Essex 2017	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	10	58%
Haines 2017	Yes	Partial	Yes	No	Yes	Partial	Yes	Yes	Yes	Yes	Yes	Yes	10	unclear
Henchoz 2010	Yes	Partial	Yes	No	Yes	Yes	Yes	Yes	Yes	No	No	Yes	8.5	57-73%
Hollinghurst 2008	Yes	No	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	10	62%
Hollinghurst 2013	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	12	56%
Jenson 2017	Yes	Partial	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	9.5	43%
Jowett 2013	Yes	No	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	10	88%
Lamb 2010	Yes	No	Yes	No	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	9	70%
Lambeek 2010	Yes	Partial	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	10.5	87%
Marra 2014	Yes	No	Yes	No	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	9	86%
McKenna 2009	Yes	No	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	10	77%
Pinto 2013	Yes	No	Yes	No	Yes	Yes	Unclear	Unclear	Yes	Yes	Yes	Yes	8	85%
Saha 2018	Yes	Yes	Yes	No	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	10	91%
Smeets 2009	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	11	79.70%
van de Roer 2008	Yes	Partial	Yes	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	9.5	89%
Vermeulen 2013	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	11	71%
Werner 2016	No	No	No	No	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	7	68.10%
Whitehurst 2015	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	11	59%
Criteria Total	**21**	**7**	**20**	**1**	**22**	**20**	**17**	**19**	**22**	**19**	**21**	**22**		

QUALITY OF COSTING METHODOLOGY IN INCLUDED STUDIES

The results of quality assessment of costing methodology are summarised in **Table 4-3**, following criteria set out by Graves et al (2002).

COSTING CRITERIA

Category 1: General costing issues (Questions 1-4 (Q1-4))

The majority of studies stated the economic perspective (Q1) (21 out of 22) with a mixture of healthcare perspective (n=4), societal perspective with disaggregated costs (n=9), both healthcare and societal perspective stated (n=8), and one with no perspective stated but with disaggregated healthcare costs. The majority of studies did not fully justify their perspective (Q2) but simply stated it within their introduction or methods (7/22 fully justified stated perspective, and 7/22 partially justified). Studies scored as 'partial' for Q2 did not give an explicit statement but gave some form of justification within the introduction/methods for their approach such as discussing at length the societal burden of the musculoskeletal condition then adopting a societal perspective. All studies apart from one study (Jenson et al, 2017) (and the study without a stated perspective (Werner at al, 2016)) gave cost data that satisfied the stated perspective (Q3).

Category 2: Methods used to determine the quantity of resources used (Questions 5-7 (Q5-7)).

As within the review by Graves et al (2002), studies rarely made a distinction between short and long run costs (Q4) (only 1/22 studies (Hollinghurst et al, 2013)). Methods for quantities of resources used by patients (Q5) were largely based on patient self-report cost diaries either collected through a questionnaire or through interview that detailed resource utilisation such as clinical visits to; GP, practice nurse, physiotherapist, hospital outpatient appointment, day case and other hospital admissions, A&E, and prescriptions (Essex et al, 2017) (21/22 studies used some form of self-report questionnaire). Other methods included using clinician inputted data (Hollinghurst et al, 2013) and extracting data from medical databases (Jenson et al, 2017,

Bultmann et al, 2009). The study by Pinto et al (2013) used the Osteoarthritis Cost and Consequences Questionnaire (OCC-Q) to collect cost data from patients, and then cross checked this information against national GP and health information databases. All studies met this criterion.

The majority of studies (20/22) described their methods for allocating time for human resources (Q6). Included studies mainly used the national average times for each type of healthcare consultation to estimate costs, other studies split time for group sessions between participants (Lamb et al, 2010, Smeets et al, 2009), and some were unclear (Bultmann et al, 2009). Methods for allocating prices between patients (Q7) were also often only clear within intervention group sessions where the cost was split between patients (Smeets et al, 2009), or through splitting the cost of intervention training (McKenna et al, 2009), although some studies detailed splitting the cost for shared capital costs such as staff travel and space provided by the NHS (Lamb et al, 2010).

Category 3: Methods used to determine the value of resources consumed (Questions 8-9 (Q8-9))

The majority of studies (19/22) gave methods for the estimation of prices, unit costs or charges (Q8), with most studies counting all resource inputs for each patient such as visits to the GP or Physiotherapist and multiplying each by the unit cost to provide the direct cost for each group of patients and then calculating the mean resource cost per patient. Published estimates of unit costs were used by the majority of studies, including for example UK NHS Reference Costs and Unit Costs of Health and Social Care publications by Curtis and Burns (2017) for secondary care and primary care costs respectively (Chuang et al, 2012, Essex et al, 2017, Hollinghurst et al, 2008, Hollinghurst et al, 2013, Jowett et al, 2013, Whitehurst et al, 2015) the Danish National Health Insurance Service Register (primary care), Danish National Patient Register (secondary care) and Danish National Prescription Registry (medication) (Jenson et al, 2017), the Dutch

Guidelines or Dutch Central Organisation for Healthcare Charges (primary and secondary care costs) or the Royal Dutch Society of Pharmacy (medication) (Bosmans et al, 2011, van der Roer et al, 2008), or private practice charges (Pinto et al, 2013, Haines and Bowles, 2017). Lamb et al (2010) stated that intervention costs were estimated using in-trial analysis but did not give reference to the full costing methods. All studies used more than just hospital charges to calculate costs (Q9), including for example medication costs in addition to hospital charges (Bultmann et al, 2009, Haines and Bowles, 2017), or using a variety of costing resources.

Category 4: Reporting of data (Questions 10-12 (Q10-12))

The majority of studies reported the year cost data were collected (Q10), with only 3 studies omitting this level of detail (Essex et al, 2017, Henchoz et al, 2010, van der Roer et al, 2008). The base cost year was also reported for most studies (Q11) with this information not present in only one study (Henchoz et al, 2010). Adjustments for costs made in different time periods (Q12) were frequently not necessary due to the majority of studies being run over a 12 month period and therefore not needing to discount for different time periods, with the exception of Barton et al (2009), who discounted at a rate of 3.5%. The majority of other studies stated adjusting prices for inflation to match base cost year (Jenson et al, 2017, Hollinghurst et al, 2013, Lamb et al, 2010, Marra et al, 2014, Pinto et al, 2013, Whitehurst et al, 2015).

COST DRIVERS

Detail of cost drivers extracted across studies included; highest resource unit cost, highest resource utilisation, and highest mean resource cost per patient, where available. There was minimal detail however within some studies, for example, Barton et al (2009) had a key cost driver of 'visit costs' but this was not broken down between different types of clinicians. In contrast, detailed visit costs were broken down and provided in the study by Hollinghurst et al (2013), providing detail of; hospital care/A&E visit costs, GP home visit costs, GP surgery visit

costs, GP out of hours costs, GP telephone consultation costs, district nurse home visit costs, practice nurse consultation costs, practice nurse telephone consultation costs, and healthcare assistant/phlebotomist costs.

Unit Costs: The highest unit costs were consistently found to be outpatient visits/medical specialist visits, hospital admissions (day case and hospital stay), and imaging (MRI/CT). This demonstrates that the secondary care costs make up the highest unit costs within MSK healthcare. (See **Table 4-4** for further detail and drivers)

Resource Utilisation: The highest resource utilisation across studies was found to be; primary care visits to the GP, outpatient/medical specialist visits, and physiotherapy visits (See **Table 4-5** for further detail and drivers).

Mean Resource Use Cost Per Patient: 13 studies gave data on the mean resource use cost per patient allowing for analysis of key drivers of costs within studies. Analysis of these studies showed that important drivers of MSK healthcare costs (starting with highest costs) were: primary care visits to the GP, outpatient/medical specialist visits, and physiotherapy visits. This demonstrates that predominantly high levels of utilisation of key resources were driving cost, with the top 3 drivers exactly matching those for highest utilisation (see **Table 4-6** (this table is also available with costs converted/inflated to GDP 2018 in **Appendix 4-2,** support provided by health economist (JH))**.** Converted costs however need to be viewed with caution due to the significant heterogeneity between studies including heterogeneity in MSK conditions treated, health systems, time horizons, and costing methods. Costs therefore cannot be directly compared and have instead been ranked within individual studies in order to identify the highest cost drivers.

Table 4-4 Highest unit costs

			Primary Care & Community				Secondary Care				Imaging		Private
Author	**Currency**	**Cost year**	**GP**	**Chiro/ Osteopath**	**Psychologist**	**Occ health Physician**	**Outpatient visit**	**A&E**	**Hospital stay (day)**	**Hospital admission**	**MRI**	**CT**	**Private other (PT/Acup /MT)**
Barton 2009													
Bosmans 2011	Euros	2004	20.44				56.66				178.51	105	
Bultmann 2009													
Chuang 2012	£ Sterling	2008-2009	45				41	123		2515			55
Essex 2017	£ Sterling	2012-2013	34				108	115	693	1877.86			
Haines 2017													
Henchoz 2010													
Hollinghurst 2008	£ Sterling	2005	24				133	61		2065			71.2
Hollinghurst 2013	£ Sterling	2009	27				103						
Jenson 2017	Euros	2015											
Jowett 2013													
Lamb 2010													
Lambeek 2010	£ Sterling	2007	16		59	162	53			261			
Marra 2014	$ Canadian	2009	34	120			99				895		67.33
McKenna 2009	£ Sterling	2005-2006	20.75				90	110	90	188			70
Pinto 2013	£ Sterling	2009	62	47			158	322	650	13576	836		
Saha 2018	Euros	2013	366		415		424						
Smeets 2009	Euros	2003	20.2		60.5	20.2	56			337			
van de Roer 2008	Euros	2004	20.4	42.5			56.6			341	223	128	
Vermeulen 2013	Euros	2008	22				74			439	179	147	
Werner 2016	$ US	2012	98	86			216			28323			
Whitehurst 2015	£ Sterling		31	38			124		562	1157	179	100	
Total in top 4			**3**	**2**	**3**	**3**	**11**	**4**	**4**	**11**	**6**	**3**	**2**

Top 2 (1st/2nd) unit costs within studies

Next 2 (3rd/4th) top unit costs within studies

* Cost drivers only included within this table if they were within the top 4 unit costs within at least 2 studies

**Costs are in different currencies and cost years so are not directly comparable

***Chiro; Chiropractor, Occ: Occupational

**** Mean costs per patient reported for control/usual care/reference group

Table 4-5 Highest utilisation

Utilisation within usual care/control															
	Primary Care & Community							Secondary Care	Imaging/Investigations		Medication	Private			
Author	GP	Practice Nurse	Physio-therapist	Manual Therapy	Other HCP	Psychol-ogist	All primary care	Outpatient visit	Xray	All invest-igations	Prescription meds	Private consults	Private other HCP		
Barton 2009	0.15	0.41			0.2			0.28							
Bosmans 2011	0.6		1	0.3				0.05	0.03	0.05					
Bultmann 2009	69		66		21	3									
Chuang 2012															
Essex 2017	3.59	1.06	1.12					0.77			3.83	2.11	0.14		
Haines 2017															
Henchoz 2010	0.5		2.3		0.1			1.1							
Hollinghurst 2008							0.43	0.32			0.85				
Hollinghurst 2013	0.77	0.04			0.02			0.17			1.36				
Jenson 2017															
Jowett 2013					3.13		2.75	0.58		0.26		0.03	0.41		
Lamb 2010	1.86	0.11	0.9			0.09		0.21	0.28	0.83	5.33	0.06	1.79		
Lambeek 2010	0.6		21.7	5.4		0.9		0.6		4.7			2.8	Physio private	
Marra 2014	0.67							0.09	0.62				2.1		
McKenna 2009															
Pinto 2013	1.75	0.12	0.75		0.12			0.49	0.35						
Saha 2018															
Smeets 2009	2.12		7.36			0.34		1.55	0.26						
van de Roer 2008	1.4		2.1	0.2				0.2	0.3						
Vermeulen 2013															
Werner 2016															
Whitehurst 2015	1.32	0.13	1.45					0.3	0.12			0.18			
Total in top 4	**11**	**1**	**9**	**2**	**3**	**1**	**2**	**8**	**3**	**1**	**4**	**2**	**4**		
		Top 2 (1st/2nd) mean utilisation within studies													
		Next 2 (3rd/4th) top mean utlisation within studies													
	* Cost drivers only included within this table if they were within the top 4 mean utilisation within at least 1 study														
	** Utilisation reported for control/usual care/reference group														
	*** HCP; Health Care Professional, meds; medications														
	****Values are mean resource use per patient														

Table 4-6 Highest mean cost per patient

Mean cost per patient																
		Primary Care & Community						Secondary Care			Imaging/Investigations		Medication	Private		
Author	Currency	GP	Physio-therapy	Manual Therapy	Psychologist	Complementary Medicine/ET/ Massage	Other HCP	Outpatient visit	Hospital stay day case	Hospital admission	Xray	MRI	Prescription medicine	Private other HCP	Equipment /aids	
Barton 2009																
Bosmans 2011	Euros	10.22	23.02	10.36				2.83			1.25	3.57	10.00			
Bultmann 2009																
Chuang 2012																
Essex 2017	£	122.06	40.51			4.90		83.16	62.37	56.34			32.06			
Haines 2017																
Henchoz 2010	Euros	31.7	73.1				12.2	106.4								
Hollinghurst 2008	£	11.03					0.35						9.82	169.65		
Hollinghurst 2013	£	19.21	69.73				0.03	30.74		51.02			11.04			
Jenson 2017	Euros	306	464				43	394		263			17	616		
Jowett 2013	£						117.1						2.54	16.17		
Lamb 2010																
Lambeek 2010	£															
Marra 2014	$ (Can)	22.59				24.58		10.93			21.45	40.68		12.24	89.74	
McKenna 2009	£	83	26				10	53						77	96	
Pinto 2013																
Saha 2018	Euros	255.77	293		106.89		64.89	65.89								
Smeets 2009	Euros	42.82	168.54		20.57		2.42	86.8		29.95			58.04			
van de Roer 2008	Euros	28.56	48.3	6.44		4.66		11.32			14.94		13			
Vermeulen 2013	Euros	23.7	253.8	87		16.5	36.5	126.9		88.9	39.8	53.3	227.7			
Werner 2016																
Whitehurst 2015	£	40.85	34.19					34.88		16.53	3.69	20.65	19.2	55.09		
Total in top 4		**11**	**9**	**1**	**1**	**1**	**1**	**9**	**1**	**3**	**1**	**1**	**4**	**3**	**2**	
		Top 2 (1st/2nd) mean costs within studies														
		Next 2 (3rd/4th) top mean costs within studies														
	* Cost drivers only included within this table if they were within the top 4 mean costs within at least 1 study															
	** Mean costs reported for usual care/control/reference group															
	***HCP; Health Care Professional, ET; Exercise Therapy															

Discussion

This review captured MSK economic analyses across a number of varied health systems/nations including the UK, the Netherlands, Denmark, Norway, Switzerland, Sweden, Italy, the US, Canada, Australia and New Zealand, and focussed on trials providing treatment primarily within primary and community healthcare settings.

QUALITY OF COSTING METHODS

Only one study (Hollinghurst, 2013) satisfied all of the costing criteria reported by Graves et al (2002). The lowest conforming study satisfied 7 of the 12 costing criteria (58%) (Werner et al, 2016), which could imply errors in costing methods (Graves et al, 2002). 14/22 (64%) of studies however satisfied at least 10 or more criteria.

In the review by Graves in 2002, only 2 out of the 12 costing quality questions (Q5 and Q8) were satisfied by more than 67% of included articles. In this review 10 of the 12 questions (all except Q2 and Q4) were satisfied by more than 77% of included articles.

The quality assessment of costing methods within this review demonstrates that more attention still needs to be given to the costing methods used to estimate individual patient costs with particular attention needed to clarify short and long run costs and to fully justify the chosen perspective. There appears however to have been a significant improvement with costing within economic analyses when these results are compared to the previous similar review of 45 economic analyses by Graves in 2002.

KEY COST DRIVERS

The highest resource unit costs were secondary care based costs, in contrast to the highest resource utilisation, which included primary, community and secondary healthcare drivers such as; GP visits, Physiotherapy visits, Outpatient/Medical Specialist visits and Prescriptions. When these were evaluated together to give the mean resource use cost per patient the same cost

drivers for utilisation were driving mean resource use costs per patient, showing that the high level of utilisation rather than the initial unit cost seems to be the predominant factor in driving mean costs within MSK healthcare in this setting (See **Tables 4-4, 4-5,** and **4-6**). These findings therefore support a shift in focus for those commissioning and evaluating MSK services, from high cost procedures (such as surgical interventions and inpatient stays) within MSK healthcare (largely at a secondary care level) to high utilisation of key resources such as GP visits and Physiotherapy visits within primary care and community settings, with further scrutiny of the cost effectiveness of the entire pathway of MSK healthcare provision.

We summarised the key drivers of MSK healthcare costs within an international context and found that the highest mean costs across studies (in order) were for: GP visits, Physiotherapy visits, Outpatient Medical Specialist visits, Prescription Medication and Hospital Admissions (day case and elective stay grouped together). These 5 drivers captured over 70% of the costs in the majority of studies with fully disaggregated costs (7/11 studies). Recommendations from the review would be to collect detailed costs for the above key drivers particularly in regards to GP, Physiotherapy and Outpatient visits which formed over 50% of the costs across the majority of studies (8/11 of those with fully disaggregated costs) (these 3 drivers alone captured over 75% of costs in two recent studies (Jenson et al. 2017, Saha et al, 2018). This detail might include capturing standard treatment times within clinics (such as 30 minute or 45 minute Physiotherapy consultations) and grade of treating clinician (basic grade or advanced level) in order to further improve the accuracy of costing for the most important cost drivers. Hollingsworth et al (2013) was the only study to detail the treating clinician grade to further improve the accuracy of their costing calculation and overall provides a useful exemplar of high-quality costing methods within community MSK healthcare. Grieve et al (2010) supports this individualised approach, highlighting that cost effectiveness analyses using average unit costs can report inaccurate incremental costs. Clearly some studies were restricted by the research costs and practicalities of collecting this level of data, however, capturing this detail would improve the accuracy of costing

information in this setting as these are key cost drivers that are not standardised across MSK services, as was too often assumed. This is particularly relevant as significant system changes are made within MSK healthcare such as the introduction of First Contact Practitioners (FCP) in the UK, where GPs are being replaced in certain areas by advanced MSK Physiotherapists to assess and manage MSK patients in primary care (NHS England and NHS Improvement, 2019). Using a standardised mean unit cost for the Physiotherapist contact in these instances would not take into account the seniority of clinicians used or the differing consultation times allocated and could therefore lead to inaccurate evaluations of cost.

Additional costs that are useful to collect dependent on the perspective of the economic review are; private healthcare professional costs (such as private Physiotherapy, Acupuncture, Osteopathy), equipment (including patient aids, orthotics etc) and imaging costs (MRI/Xray were the most important of these).

LIMITATIONS OF THE REVIEW

Limitations of this review were that not all MSK economic analyses were included within the review, which means that our results cannot be generalised to all MSK economic analyses. Non-English studies were also not included due to the resource limits of this review. Only direct healthcare costs were reviewed, future studies incorporating additional societal costs would be useful to further inform population level health systems. This study was a broad review of MSK economic analyses and due to this breadth, there was a large amount of heterogeneity between studies. This means that caution needs to be applied when looking to make direct comparisons between studies. Future research in this area focusing on one health system or one condition would allow for a more detailed and in-depth analysis of direct costs. Further review of other sources of activity/spend data alongside prospective studies evaluating the cost drivers identified within this review in primary/community MSK services would also be useful to further explore and validate findings.

Conclusion

This review provides a detailed overview of the quality of costing methods used within MSK economic analyses and has identified key drivers of MSK healthcare costs for patients accessing treatment in community and primary care settings. The quality and accuracy of costing data in this setting needs more attention around capturing the grade of treating clinicians and specific consultation length for clinical visits to more accurately determine true patient-level costs as these factors were key cost drivers. If this information was collected in a standardised, accurate and consistent manor, it could form a useful part of a standardised MSK dataset (alongside key metrics measuring treatment outcome/performance) and help to develop future benchmarking capabilities within these settings supporting national data evaluations and informing healthcare policy (such as NHS RightCare (2019) in the UK setting). Such an approach would also support a future direction towards value-based care (health outcomes achieved per monetary unit spent) which looks to achieve good outcomes in the most efficient way (Porter, 2009), helping to form a system level framework for restructuring healthcare delivery for the future.

Summary of Chapter

This chapter has reviewed NHS payment practices and costing methods within MSK economic analyses and identified key cost drivers and existing costing tools used within MSK community and primary care.

Findings from the review of NHS payment practices highlights the prioritisation of patient level costing (PLICS) moving forwards (NHS England and Improvement, 2020), with NHS Improvement supporting the adoption of this new costing process across acute and community services. Following the review of NHS costing methods, a systematic review of economic analyses within the field of MSK was undertaken with a view to identifying costing methods and important cost drivers within MSK community and primary care services. When taking the findings from the review of economic analyses and viewing and analysing these alongside details set out for the new PLICS costing process, as expected there is significant overlap. The PLICS costing system is complex and made up of multiple activity codes and unique patient and clinician identifiers. This data will be extracted from clinical information systems and digital platforms and pulled into business unit dashboards to allow for costing (NHS England and NHS Improvement, 2020). Areas highlighted by the economic evaluation are captured within the PLICS system but in a slightly different format. For instance, there are codes for point of delivery (POD) which include; emergency attendance (EA), outpatient attendance consultant led (CL), outpatient attendance non-consultant led (NCL), outpatient procedure (OPROC), elective inpatient (EL), non-elective short stay (NES) and non-elective long stay (NEL) that would capture contacts throughout the patient pathway of MSK care. PLICS also details; care team referred to, contact date, clinical contact duration, consultation type, attendance code (new/follow up, face to face/non face to face), activity location, group therapy indicator, attendance status, care professional, and onward referrals. The PLICS system is not however yet fully embedded and due to its complexity would benefit from being condensed into a more useable format in the short-term to prioritise the most important data items/cost drivers that could be compared across MSK services. If successfully

embedded however in the future, it would allow for the detailed collection of the key cost drivers identified within the systematic review.

Themes arising from the economic review include the importance of capturing clinical visits (GP/Physiotherapy/Outpatient Medical Specialist) but did not often include the type and grade of treating clinician or specific visit times. This information will be captured within the NHS PLICS system in the future. Applying methods for estimating time of clinical contacts ideally with individual services able to input contact times directly, or by using standard service contact times such as a specific First Contact Practitioner (FCP) service providing 20-minute appointments with a Physiotherapist, or using national averages such as GP appointments at the average 9.22 minutes (Hobbs et al, 2016) where this is standard practice would further enhance the accuracy of costing tools. This enables accurate unit costs to be calculated for comparison of models. Other areas include the need to capture the costs of the full pathway of care and not limit costing to the care provided in community and primary care. This is demonstrated by the point of delivery code within the PLICS system and is highlighted within economic analyses as hospital admissions, investigations, and outpatient medical specialist appointments. The importance of understanding the impact different MSK services or models of care have on investigation rates, orthopaedic outpatient consultations, inpatient admissions, and prescribing practices is demonstrated by both the economic analysis identifying these variables as important drivers of MSK healthcare costs, and the review of existing costing tools and guidance that looks to demonstrate the improved efficiencies of new models of care through reduced use of these high cost drivers (HEE, 2020).

Lord Carter detailed the level of efficiency savings that could be realised with changes in the way that NHS services are delivered and evaluated, learning from those services deemed to be achieving high value healthcare for patients (Carter, 2016). It is clear that in order to fully evaluate and compare the costs associated with community and primary care MSK services that a

number of resource variables that drive the majority of costs within the MSK pathway of care need to be captured accurately and consistently to allow for transparent and useful comparison.

The next step following this in-depth review of costs, was to review the identified resource variables that appear to be driving the majority of MSK costs with a group of MSK Stakeholders (**Chapter 5**) with knowledge of data collection processes within the NHS in order to agree a realistic and detailed list of data items for routine collection within MSK services to allow for direct comparison. Cost data was considered alongside data items reviewed in previous chapters (for case-mix adjustment). Together these data items alongside PROMs/PREMs could allow for a full benchmarking methodology to be developed and tested across MSK community and primary care services paralleling existing approaches by the National PROMs Programme (NHS England, 2016), and NHS RightCare (2016).

5 Chapter 5: Developing a core outcome set for community and primary care musculoskeletal services: a consensus approach.

Previous chapters have explored metrics commonly used for case-mix adjustment of MSK PROMs (Burgess et al, 2019, **Chapter 2**), for predicting MSK functional outcome (Burgess et al, 2020a, **Chapter 3**), and for capturing key MSK costs (Burgess et al, 2020b, **Chapter 4**). This Chapter takes the findings of these reviews alongside other national data collections to develop a list of proposed MSK metrics for routine collection in community/primary care MSK services. This proposed list was then taken to consensus to better understand acceptance, feasibility, and barriers to moving towards a standardised data collection, and agree a minimum list of metrics/outcome measures to form a core outcome set (COS) to be recommended for routine collection across MSK community and primary care services.

Abstract:

Background: At present there is no core outcome set (COS) for use in community and primary care Musculoskeletal (MSK) services across the UK to the authors knowledge. Services are therefore collecting different MSK outcomes and metrics in different ways and at different times. Standardising MSK data collection is essential in order to allow for fair and impactful benchmarking to be undertaken, and in order to facilitate improvements in the quality of care delivered to the millions of patients presenting each year with MSK disorders.

Objective: To gain consensus on a proposed set of metrics that could be used to develop a core outcome set for use in routine practice in community and primary care MSK services in the UK and to make recommendations to inform a future national MSK audit.

Methods: A consensus process involving researchers, healthcare professionals and patients. Previous research generated an initial list of proposed metrics. This proposal was then taken to

wider stakeholder consensus via an online survey designed for both healthcare professionals and MSK service users.

Results: 199 respondents completed the survey, 166 healthcare professionals and 33 service users (25/33 eligible to answer all items within the survey). Metrics that reached strong consensus were; age, pain site, comorbidities, duration of symptoms, work status, work absence, work absence duration. No Patient Reported Outcome Measures (PROMs) met strong consensus and all Patient Reported Experience Measures (PREMs) other than timeliness/convenience met strong consensus criteria.

Conclusion: 7 baseline factors and 9 PREM domains reached strong consensus. The MSK-HQ PROM was the highest rated outcome measure so was also recommended for inclusion in an MSK core outcome set.

Contribution to the Thesis

- There is strong consensus that standardising MSK data would improve patient care
- Most patients surveyed would be happy to fill in pre/post treatment questionnaires
- The MSK-HQ was the preferred generic PROM
- Nine PREM domains reached strong consensus
- Age, pain site, comorbidities, duration of symptoms, and work metrics are important

The contents of this chapter have been published in part in:

Burgess, R., Lewis, M., McRobert, C. and Hill, J.C., 2021. Developing a Core Outcome Set for Community and Primary Care Musculoskeletal Services: A Consensus Approach. *Musculoskeletal Science and Practice*, p.102415

Introduction

The NHS Mandate (2019) lays out the need for greater transparency about the quality of NHS care, with a focus on reducing inequalities and unwarranted variation in care provision and patient outcomes. The Berwick Report (2013), written to ensure learning was realised from known failures in patient safety and NHS care concluded that;

'*The most important single change in the NHS in response to this report would be for it to become, more than ever before, a system devoted to continual learning and improvement of patient care, top to bottom and end to end.*'

In order to do this, healthcare systems need to offer standardised, efficient data entry, and allow for collection and collation of system, clinical, and patient-reported outcomes. This will help to monitor and address the escalating burden of MSK pain conditions, and help to plan, monitor, and improve the quality of MSK healthcare provision, through continual evaluation and scrutiny, and sharing of best practice exemplars (Slater and Briggs, 2017).

Arthritis Research UK published a specific MSK Recommended Indicator Set in 2016 (ARUK, 2016). A number of national audits are already in place in the UK to capture MSK data in specific areas outlined within the indicator set, such as; the National Early Inflammatory Arthritis Audit (NEIAA) (BSR, 2019), the National Hip Fracture Database (NHFD) (RCP, 2019), the National Joint Registry (NJR, 2020) and the National PROMs Programme hip and knee data (NHS England, 2020). The area however where there is no national audit or dataset to provide comparative data is within community and primary care management of non-inflammatory, non-surgical MSK conditions. For these non-surgical, non-inflammatory patients at present, we do not have continuous national data to support the indicators listed under 'Musculoskeletal Health Outcomes' within the Versus Arthritis Indicator set. This includes data on health gain (MSK-HQ), or utility scores (EQ5D), or other key metrics such as the percentage of patients whose work is affected by MSK

conditions, and the satisfaction of patients treated for MSK conditions in this setting (ARUK, 2016).

A recent evaluation of First Contact Physiotherapy services in primary care commissioned by the Chartered Society of Physiotherapy (CSP) was a first step towards a national audit of MSK primary care services (CSP, 2020). Developing methodology for routine continuous collection of MSK metrics however was not an objective of the project. Findings from this evaluation therefore need to be taken alongside additional research studies and embedded routine data collections to develop this, in order to allow for a successful benchmarking/audit tool for the future.

A number of core outcome sets (COS) exist for specific MSK conditions and are listed on the COMET database (COMET, 2020). No COS however to the author's knowledge exists encompassing all MSK conditions for use in routine clinical practice within a community or primary care setting. This study therefore aims to gain consensus across the UK on the standardised data that would need to be included in a community and primary care MSK wide national data collection. Results will be used to develop recommendations to inform a future national MSK audit.

Metrics for consideration within this COS will include variables needed to: 1. Case-mix adjust data in order to make fair comparisons between outcomes achieved (Burgess et al, 2019, 2020a **Chapter 2-3**); 2. Describe the population in order to identify unwarranted variation between patient groups (e.g. age, ethnicity, gender); 3. Identify treatment effect including use of Patient Reported Outcome Measures (PROMs); 4. Capture patient satisfaction including use of Patient Reported Experience Measures (PREMs); 5. Monitor effective implementation of best practice recommendations including NICE guidelines and other MSK Indicators; and 6. Allow for costing of MSK services using key MSK cost drivers (Burgess et al, 2020b, **Chapter 4**).

Objectives:

Primary Objective: To gain consensus on a proposed set of metrics that could be used to develop a COS for use in routine practice in community and primary care MSK services in the UK. This dataset needs to be considered feasible and appropriate for collection by clinicians in clinical systems, and feasible and appropriate to patients who will provide the majority of the data in the form of questionnaires to be collected and collated by individual MSK services.

Secondary Objective: The secondary objective is to aid development of methodology for a national audit of community and primary care MSK services.

Methods

ETHICAL APPROVAL:

Ethical approval for the study was granted by the University's Faculty of Medicine and Health Sciences Research Ethics Committee (REC reference: MH-200141) (see **Appendix 5-1**)

GUIDELINES:

The MSK COS development followed recommendations set out by the COS-STAD team (Kirkham et al, 2017)[1]

The MSK COS development was also listed a priori on the COMET (Core Outcomes Sets for Effectiveness Trials) database.[2]

CONSENSUS OVERVIEW

A consensus process involving researchers, healthcare professionals and patients was conducted, starting with multiple reviews of related literature by the research team to generate an initial list

[1] Kirkham et al, 2017; COS-STAD: https://doi.org/10.1371/journal.pmed.1002447.t002
[2] COMET database; link to study: http://www.comet-initiative.org/Studies/Details/1551

of proposed metrics. This proposal was then taken to wider stakeholder consensus via an online survey.

Domain 1: Scope

The key objective of the MSK COS development is to provide clear guidance to the MSK community on which generic (MSK wide) metrics and outcomes should be collected as a minimum in routine clinical practice across the UK (see **Table 5-1** for detail on scope).

Table 5-1: Scope, stakeholders, and consensus process, following COS-STAD standards for development

Standard Number	Domain	Methodology (COS-STAD)	Detail
1	**Scope**	**The research or practice setting(s) in which the COS is to be applied**	UK Community and Primary Care setting for use in routine clinical practice
2	**Scope**	**The health condition(s) covered by the COS**	All MSK conditions (focus on non-inflammatory, non-surgical)
3	**Scope**	**The population(s) covered by the COS**	All adult (18 years or over) MSK patients attending for routine MSK care.
4	**Scope**	**The intervention(s) covered by the COS**	All interventions (focus on non-surgical)
5	**Stakeholders involved**	**Those who will use the COS in research**	Researcher involvement: The MSK research team at Keele developed the preliminary recommendations based on best evidence and expertise in this area.
6	**Stakeholders involved**	**Healthcare professionals with experience of patients with the condition**	Survey content specifically designed for healthcare professionals (including physiotherapists, GPs, managers, commissioners and other professionals working within/with MSK community and primary care services) captured views from healthcare professionals.
7	**Stakeholders involved**	**Patients with the condition or their**	Survey content specifically designed for patients

		Representatives	gained feedback on outcome measures and core metrics for potential use in routine clinical practice.
8	**Consensus process**	**The initial list of outcomes considered both healthcare professionals' and patients' views.**	The survey will have 2 arms, one for healthcare professionals and one specifically to capture patient's views.
9	**Consensus process**	**A scoring process and consensus definition were described a priori.**	The scoring process and consensus definition was developed a priori and detailed within the original protocol submitted for ethical approval and outlined below.
10	**Consensus process**	**Criteria for including/dropping/adding outcomes were described a priori.**	The consensus aimed to reduce and not add outcomes/metrics where possible aiming for a feasible list for routine capture. Criteria for including/dropping metrics were described a priori within the protocol and are outlined below.
11	**Consensus process**	**Care was taken to avoid ambiguity of language used in the list of outcomes.**	Supporting documentation giving full detail of tools/outcome measures/metrics and abbreviations were included and easily accessible from the survey webpage for all healthcare professionals. The patient survey included video clips to explain all aspects of the proposed survey.

This Table follows recommendations for reporting by the COS-STAD study team (Kirkham et al, 2017)

Domain 2: Stakeholders

The stakeholders involved in developing and providing feedback for the consensus study are outlined in **Table 5-1**. Inclusion of researchers, healthcare professionals and patients is required as a minimum by the COS-STAD recommendations (Kirkham et al, 2017)**.**

The aim was to recruit a minimum of 100 healthcare professionals (clinicians, managers, MSK programme leads) including representation from across the UK and across professional groups. The study also aimed to recruit a minimum of 25 patients in order to gain essential feedback on feasibility and acceptability of a proposed patient questionnaire to patients using MSK services. This was a pragmatic choice like many similar consensus studies with the main aim of ensuring representation across stakeholder groups (Williamson et al, 2017).

Domain 3: Consensus Process

The consensus process was developed following standards set out by Kirkham et al (Kirkham et al, 2017) (see **Table 5-1**)

CONSENSUS STAGE 1: PRELIMINARY COS RECOMMENDATIONS

Three published reviews have been undertaken by the research team to develop preliminary recommendations on the content of the MSK COS. These include; a systematic review of MSK case-mix adjustment models and an umbrella review of predictors of MSK outcome to identify variables necessary to adjust MSK data in order to make fair comparisons (Burgess et al, 2019, 2020a, **Chapters 2-3**), and a review of economic analyses in MSK community and primary care to identify key cost drivers of MSK healthcare provision (Burgess et al, 2020b, **Chapter 4**). Additional national recommendations and guidance and other identified condition specific COS have also been reviewed to form an initial proposal that formed the basis for the consensus process (ARUK, 2016, NHS England 2020, CSP, 2020, NICE, 2016, COMET, 2020, ICHOM, 2020a 2020b, NHS England, 2019a 2019b). Evidence supporting inclusion of listed metrics is detailed below.

Evidence synthesis:

Within the standards developed by Kirkham et al (2017) on COS development, consensus from stakeholders was not reached on the need for a systematic review to identify outcomes/metrics to be considered within the consensus process. For this COS development however, two

systematic reviews and one umbrella review led by the MSK COS development team on case-mix adjustment (Burgess et al, 2019, **Chapter 2**), key drivers of MSK healthcare costs (Burgess et al, 2020b, **Chapter 4**), and predictors of functional outcome (Burgess et al, 2020a, **Chapter 3**) were used to develop an initial recommendation alongside a wider review of the evidence and latest national guidelines. The COMET database was also searched to ensure that no other generic MSK COS for use in the community/primary care UK setting existed and also to identify other MSK outcome sets that could help inform/underpin recommendations. National guidelines, recommendations and evaluations were also used to inform the final proposed list and included; the Versus Arthritis recommended indicator set (ARUK, 2016), NICE guidelines (NICE, 2016), and the national FCP evaluation (FCP, 2020).

Developing a proposed dataset:

The proposed dataset is made up of core areas of; demographic factors, clinical factors, employment factors, functional/MSK health status, patient reported experience measures, and healthcare utilisation (economic factors). This is a collection of evidence based validated tools and patient factors/metrics including demographics and characteristics that can be used for case-mix adjustment. All proposed mandatory tools included are free to use subject to obtaining the associated licence agreements.

Case-mix adjustment: Metrics supported for inclusion by the systematic review on case-mix adjustment (Burgess et al, 2019, **Chapter 2**) and the umbrella review of predictors of MSK outcome (Burgess et al, 2020a, **Chapter 3**) are shown in **Table 3-8** (**Chapter 3, page 110**) and are discussed in detail at the end of **Chapter 3**.

Cost Metrics: The key drivers of MSK healthcare costs identified in **Chapter 4** (Burgess et al, 2020b) were found to be (in order): GP visits, Physiotherapy visits, Outpatient Medical Specialist visits, prescription medication and hospital admissions (day case and elective stay grouped together). Additional costs that are useful to collect dependent on the perspective of the

economic review are; private healthcare professional costs, equipment, and imaging costs (MRI/Xray were the most important of these).

Health outcomes supported by the Versus Arthritis MSK Indicator Set (ARUK, 2016)

Musculoskeletal Health Outcomes:

VA Indicator 20: 'Change in health utility score from initial presentation to six–months after management (EQ–5D or Musculoskeletal Health Questionnaire: MSK–HQ).' (ARUK, 2016)

Musculoskeletal Health Questionnaire: The Musculoskeletal Health Questionnaire (MSK-HQ) was developed collaboratively between Keele University and Oxford University supported by Versus Arthritis (formally ARUK) in 2016 (Hill et al, 2016). The collaboration aimed to address the need for a generic MSK outcome measure that could be used throughout a pathway of care and across MSK conditions to help standardise, simplify and promote MSK PROM capture across MSK settings (Hill et al, 2016). The questionnaire was developed following consensus workshops between key stakeholders including patient representatives to review and identify potential outcome domains for inclusion. The essential domains agreed were; pain severity, physical function, work interference, social interference, sleep, fatigue, emotional health, physical activity, independence, understanding, confidence to self-manage and overall impact. Four validation cohorts demonstrated that the MSK-HQ had high completion rates (94%), excellent test re-test reliability (intra-class correlation coefficient =0.84) and strong convergent validity compared with reference standards (against EQ5D Spearman's rank correlation coefficient =0.81) (Hill et al, 2016).

Gibbons and Fitzpatrick (2018) conducted a case study of the introduction of the newly developed MSK-HQ into practice. Eleven partner organisations took part in the pilot study. The MSK-HQ was found to be feasible and practical for use with patients to support patient care by identifying the main presenting problems. It was also agreed that the aggregated data could be

useful for service planning and informing commissioners/decision makers overseeing MSK services. There were, however, logistical barriers in the form of collection, with most services resorting to paper-based collection due to the complexity of integrating with clinical systems and cost of more technologically advanced solutions (Gibbons and Fitzpatrick, 2018). These barriers remain a priority to overcome to increase the feasibility and sustainability of using PROMs such as the MSK-HQ in a standardised way across health systems.

A further study in 2020 investigated the responsiveness of the MSK-HQ in measuring clinical change as a single generic MSK PROM following MSK treatment (Price et al, 2020). Four cohort studies were conducted which included 592 patients with differing MSK conditions, treated in surgical or non-surgical settings (hip, knee, shoulder, general MSK (Physiotherapy cohort)). The MSK-HQ demonstrated strong correlation (R=0.73) across cohorts with the EQ5D, and with each of the joint specific Oxford PROM scores (hip R=0.87, knee R=0.92, shoulder R=0.77). Additionally, the MSK-HQ measured the greatest treatment effect in all subgroups compared with the EQ5D.

The MSK-HQ has been shown to date to be reliable, valid, feasible to collect, and highly responsive to change compared to disease specific and generic comparator measures. With the success of the measure in providing a generic solution to measuring outcome across MSK conditions and care pathways demonstrated, the next step is to evaluate whether there is widespread support from the clinical community to collect it in routine practice, and what measures need to be collected alongside it to maximise the potential to inform and evaluate MSK care.

EQ-5D: The EQ-5D is a measure of quality of life and is an important tool for economic evaluation (Euroqol, 2021). It was developed to describe and value health across a wide range of disease areas producing a single summary value for health status by measuring five dimensions; mobility, self-care, usual activities, pain/discomfort and anxiety/depression. This allows for the calculation of Quality Adjusted Life Years (QALYs) used in economic evaluations of healthcare interventions

(Euroqol, 2021). It is recommended as part of the VA (ARUK, 2016) indicator set and was previously recommended by the CSP as a generic PROM for use in routine physiotherapy care (CSP, 2016). The EQ-5D requires a licence agreement and is not free for all users (previously a licence had been provided by the CSP but this has now expired (CSP, 2016)). The EQ-5D has therefore been added to the proposed list as a potential optional PROM subject to licencing.

VA Indicator 21: 'Percent of people of working age locally who are receiving Employment Support Allowance due to a musculoskeletal problem.' (ARUK, 2016)

A number of key work-based metrics/outcomes have been included within the proposed dataset, these include work status, work absence, work absence duration and benefit status. These mimic those that were used as part of the National First Contact Practitioner (FCP) Evaluation Project led by Keele University and funded by the CSP and NHS England (CSP, 2020).

VA Indicator 22: 'Patient experience of musculoskeletal health care services.' (ARUK, 2016)

Patient experience measures (PREMs) have also been included within the proposed dataset as these are considered an essential component of MSK service and system evaluation (CSP, 2020, Fennelly et al, 2018). A list of potential patient experience metrics including validated questionnaires such as; the Friends and Family Test (NHS England, 2019) to collect information on patient satisfaction, CollaboRATE (Elwyn et al, 2013) to collect information on shared decision making, and the Valuing Patients as Individuals (VPAI) Questionnaire to gather information on care and respect and understanding and engagement (Coyle and Williams, 2001), have been included within the proposed dataset to gauge opinion on importance of inclusion of these types of PREMs. Again, these are matched closely to those used as part of the FCP national evaluation (CSP, 2020) ensuring that all tools within the proposed mandated dataset are free to use subject to licence requests being made.

Metrics Supported by NICE Guidelines

Risk Stratification: Another mechanism for describing and analysing complexity of MSK patients in community and primary care is through the use of specific risk stratification tools. The NICE guidance for low back pain (LBP) and sciatica (NICE, 2016) recommends the use of risk stratification tools such as STarT Back (Hill et al, 2011) for each new episode of LBP to inform shared decision making and management planning. An additional risk stratification tool called STarT MSK has more recently been developed by the Keele research team extending the STarT Back to include the other most common MSK pain presentations (Campbell et al, 2016, Hill et al, 2020), making it more implementable across an MSK pathway of care. This tool was captured as part of the national FCP evaluation (CSP, 2020) and has been included within proposed optional metrics.

Metrics Supported by Review of COMET Database (Core Outcome Measures in Effectiveness Trials)

The COMET website and database were launched in August 2011. The database includes studies where COS have been developed alongside studies relevant to COS development. There are over 120 COS tools listed on the database. The COMET initiative aims to encourage evidence-based COS development and to reduce variation in approaches (COMET, 2020).

Existing COS identified for use in routine MSK community/primary care practice

The search of the COMET database included searching; 'Rehabilitation', 'Orthopaedic and Trauma' and 'Rheumatology', capturing all MSK conditions, and included studies published within the last 10 years (2010-2020).

Searches of the COMET database identified 94 results for 53 COS studies. COS studies were considered for further evaluation if they were for adult patients, were identified by the developers as appropriate for use in clinical practice and were relevant to a non-inflammatory, non-surgical community and primary care patient setting. Following review of the identified

published studies 5 studies were appropriate for full text review to inform the generic MSK COS development, all of which are presented and discussed below.

Lynch et al (2013): This paper was mainly focussed to physical/objective assessment of functional outcome following ACL injury or reconstruction and although the use of PROMs was recommended, consensus was only agreed for objective functional outcomes and not for any subjective PROMs or other patient reported metrics. This paper therefore does not provide any further learning to inform a generic COS development focused to patient reported data.

Kloppenburg et al (2014): The OMERACT Hand OA Special Interest Group began the process of developing a core set of outcome measures for hand OA clinical trials, observational studies and clinical practice in 2014. Consensus was sort between 48 experts using a Delphi process. Key recommendations for COS development were; the inclusion of specific core domains of: patient global assessment, pain, physical function and joint activity (specific to OA), and the inclusion of further domains within some settings of: pain medication, quality of life, structural damage (specific to OA), reduced mobility, and reduced strength. Specific metrics were not agreed within this study but many of the domains discussed are transferable to a generic COS tool (see **Table 5-2**).

Clement et al (2015): An International group of 22 specialists came together in 2015 to review low back pain (LBP) literature and select LBP metrics to improve performance and value in the area of degenerative lumbar conditions (Clement et al, 2015). This work was supported by ICHOM, a non-profit organisation focused on the development of standard sets of outcomes and risk factors for comparative analysis and benchmarking (ICHOM, 2020). They developed an outcome set based on 6 core domains around pain, function, health related quality of life (HRQoL), work, treatment complications and medication. Case-mix (risk) adjustment was also included but not tested and included demographic variables (age, sex, socioeconomic status (SES)), baseline clinical status (Glassman Criteria, American Society of Anaesthesiology **(**ASA)

score, comorbidities, smoking status, body mass index (BMI), pain level, duration of symptoms, current analgesia), baseline functional status (Oswestry Disability Index (ODI), EQ5D, work status, sick leave duration) and previous treatments (surgery, injection therapy).

Rolfson et al (2016): A year after the LBP COS was developed (Clement et al, 2015), Rolfson et al (2016) developed a standardised set of outcome measures for patients with hip and knee OA. This work was again an international collaboration supported by ICHOM and involved 21 expert clinicians and 2 patient representatives. A modified Delphi study was undertaken which involved 8 teleconferences and voting via an online survey. Agreed domains included pain, function, health related quality of life (HRQoL), and work status. Variables included within the proposed case-mix model were; age, sex, SES (education), joint specific history, surgical history, living status, BMI, physical activity, smoking and comorbidities. This case-mix model as the model described by Clement et al (2015) above, however, has not yet been internally or externally validated within this defined population. Included domains developed as part of this COS set would be highly transferable to a generic COS and alongside the earlier paper by Clement et al, 2015 for LBP will help to inform expert consensus for a generic COS.

Klokkerud et al (2018): The aim of this study was to develop a COS to evaluate the effectiveness of multidisciplinary rehabilitation for MSK disorders. It was developed using a stepwise process of a Delphi procedure (including experts and patient representatives in Norway), a systematic literature review and a pilot study. The Delphi group agreed specific areas of health and function deemed most relevant and important to measure within this multidisciplinary rehabilitation setting. These were found to be; pain, fatigue, physical fitness, mental health, daily activities, goal attainment, motivation, quality of life, social participation, and coping. Other factors included within the pilot data collection were; age, gender, marital status, education, and employment status alongside diagnostic information. 49% of patients within the pilot study had inflammatory rheumatic conditions with the majority of patients recruited from rehabilitation institutions or

rheumatology hospital-based departments. Only 35 of the 386 patients within the pilot study were recruited from primary care outpatient clinics. This COS was also only tested in Norway and requires further validation in other countries/settings. This paper is therefore a useful paper to consider and to review priority MSK domains and specific tools recommended to measure them (although authors also acknowledge within the paper that there is a large amount of crossover between these domains and the MSK-HQ developed by Hill et al (2016) in the UK), it is however clear that this paper was focussed to a different population of patients to the proposed study. This paper focused on patients undergoing largely secondary care based rehabilitation whereas this study looks to develop a UK COS focussed like this COS to generic MSK disorders but specifically focussed to patients presenting for MSK care within community or primary care clinics.

Summary of COMET database review: Table 5-2 shows the detailed findings from the 5 identified studies. Within these studies the only PROM tools utilised by more than one study were: the numeric pain rating scale (NPRS) (called NRS-Pain within the study by Klokkerud et al, 2018) and the EQ-5D. Domains listed across more than one study included; pain, function, HRQoL, medication, and work status. All of these PROMs/domains are included within the proposed dataset.

Additional supporting evidence: A systematic review conducted in 2018 looked to identify PROMs utilised by UK Advanced Practice Physiotherapists in MSK (Fennelly et al, 2018). This study reviewed 12,302 title/abstracts and resulted in 38 studies suitable for inclusion within the review. Quality of included papers was moderate overall. Seventy two PROMs were collected across the 38 studies. This review found that the most frequently used PROMs included patient satisfaction, quality of life (QoL), functional status, and pain, and less frequently global improvement, mental well-being, work ability, and healthcare utilisation/costs (Fennelly et al, 2018). The EQ5D and SF36 were identified as the most commonly reported QoL tools, visual

analogue scales (VAS) were most used to capture pain, but there was uncertainty about how to capture functional status. Items used to capture patient experience were also not consistent with authors concluding that further exploration of the dimensions of satisfaction was warranted to provide evidence on what should be measured.

Table 5-2: Summary of Findings from Existing COS Tools Identified from COMET Database

Author (first) & Year Published	Condition	Core Domains	Case-Mix Factors	PROMs
Lynch (2013)	ACL	Effusion, giving way, muscle strength, PROM, return to sport.		None agreed
Kloppenburg (2014)	Hand OA	Patient global assessment, pain, physical function, joint activity, pain medication, HRQoL, structural damage, reduced mobility, reduced strength.		
Clement (2015)	LBP	Pain, function, HRQoL, work, treatment complications, medication.	**Demographics;** age, sex, SES. **Baseline clinical status;** Glassman Criteria, ASA score, comorbidities, smoking status, BMI, pain level, duration of symptoms, current analgesia. **Baseline functional status;** ODI, EQ5D, work status, sick leave duration. **Previous treatments;** surgery, injection therapy.	NPRS ODI EQ5D
Rolfson (2016)	Knee and Hip OA	Pain, function, HRQoL, work status.	**Demographics;** age, sex, SES. **Baseline clinical status;** joint specific history, surgical history. **Other;** living status, BMI, physical activity, smoking, comorbidities.	HOOS-PS KOOS-PS EQ5D SF12 or VR12

Klokkerud (2018)	All MSK disorders	Pain, fatigue, physical fitness, mental health, daily activities, goal attainment, motivation, quality of life, social participation, coping.	Age, gender, marital status, education, employment status.	EC-17 FFbH NRS Fatigue PSFS HSCL-5 NRS Pain EQ5D COOP/WONCA

ACL; Anterior Cruciate Ligament, BMI; Body Mass Index, COOP/WONCA; Social Participations Questionnaire, EC-17; Effective Musculoskeletal Consumer Scale 17, EQ-5D; EuroQol 5-domain instrument, FFbH; Hannover Functional Questionnaire, HSCL-5; Hopkins Symptom Checklist 5, OA; Osteoarthritis, HOOS-PS; Hip Disability and Osteoarthritis Outcome Score short version, HRQoL; Health Related Quality of Life, KOOS-PS; Knee Injury and Osteoarthritis Outcome Score short version, NRS; Numeric Rating Scale, ODI; Oswestry Disability Index, PROM; Patient Reported Outcome Measure, PSFS; Patient Specific Functional Scale, SES; Socioeconomic status, SF-12; Short Form 12 health survey, VR-12; Veterans Short Form 12 health survey.

The proposed mandatory and optional metrics are shown below in **Table 5-3** and the full explanatory document can be viewed as supplementary material (**see Appendix 5-2**).

Table 5-3: Proposed Mandatory and Optional Variables

Variable Name	Response Options	Capture Point	Justification
Mandatory Variables			
Demographics			
Age	Continuous numeric	Baseline	CM/PO review
Sex at birth	Binary (male/female)	Baseline	Descriptive
Education	Categorical (4 options)	Baseline	CM review
Ethnicity	Categorical (5 options)	Baseline	CM review
Baseline Clinical Factors			
Pain Site	Categorical (11 options)	Baseline	CM review
Comorbidities	Categorical (12 options)	Baseline	CM/PO review
Duration of Symptoms	Categorical (5 options)	Baseline	CM/PO review
Previous Surgery	Categorical (4 options)	Baseline	CM review
Self-Reported as Disabled	Binary (yes/no)	Baseline	CM review
Employment			
Work Status	Binary (yes/no)	Baseline and 3 months	VA/FCP
Work Absence	Binary (yes/no)	Baseline and 3 months	VA/FCP
Work Absence Duration	Categorical (4 options)	Baseline and 3 months	VA/FCP
Functional Status			
MSK-HQ (MSK Health Status)	Questionnaire (15 questions)	Baseline and 3 months	VA/FCP
Pain Intensity (NPRS)	Numeric (0-10)	Baseline and 3 months	PO review/COMET
Patient Reported Experience			
Friends and Family Test (FFT)	Questionnaire (2 questions)	3 months	NHSE/FCP
Global Change in Health Status	Categorical (6 options)	3 months	FCP
Optional Variables			
Baseline Clinical Factors			
Previous Physiotherapy	Binary (yes/no)	Baseline	CM review
Assisted with Questionnaire	Binary (yes/no)	Baseline	CM review
Employment			
Benefit Status	Categorical (12 options)	Baseline	CM/PO review/FCP

Functional Status			
STarT MSK (Risk Status)	Questionnaire (10 questions)	Baseline	NICE/FCP
EQ5D5L (QoL)	Questionnaire (5 questions)	Baseline and 3 months	VA/COMET
Patient Reported Experience			
Valuing Patients as Individuals -Care and Respect -Understanding & Engagement	Questionnaire (6 questions)	3 months	FCP
CollaboRATE -Shared Decision making	Questionnaire (3 questions)	3 months	NHSE
MSK Indicators -Clinical Competence -Sufficient Information	2 questions	3 months	FCP
Economic Factors			
Healthcare Utilisation	Free text numeric	3 months	HE review/FCP
Investigations and Treatments	Free text numeric	3 months	HE review/FCP
Inpatient Stays	Free text numeric	3 months	HE review/FCP
Prescribed Medication	Binary (yes/no)	3 months	HE/CM review

CM; case-mix adjustment review, COMET; core outcome measures in effectiveness trials database review, FCP; First Contact Physiotherapist National Evaluation, HE; health economics review, NHSE; NHS England, NICE; National Institute for Clinical Excellence LBP guidelines, PO; predictors of outcome review, VA; Versus Arthritis MSK Indicators (References: ARUK, 2016, NHS England 2020b, CSP, 2020, NICE, 2016, COMET, 2020, ICHOM, 2020a 2020b, NHS England, 2019a 2019b).

CONSENSUS STAGE 2: DEVELOPMENT OF ONLINE SURVEY RELATING TO CANDIDATE METRICS/OUTCOME MEASURES

Survey: A single online self-administered survey was developed using Lime Health Survey software provided by Keele University. The survey was designed to take no more than 30 minutes for health professionals or patients to complete. Respondents were signposted to the appropriate questions (healthcare professional or patient/service user) dependent on their responses to initial questions. Please see **Appendix 5-3** for full survey content. The survey was designed to gauge opinion from healthcare professionals on the importance of the proposed metrics/outcomes for inclusion in a standardised dataset and to gain information on current

practice in capturing MSK data, alongside feasibility to implement a standardised COS in clinical practice. The patient arm of the questionnaire displayed the proposed questionnaire that would be needed to capture key metrics and outcomes to patient participants. This was displayed in sections using video clips. Participants were then asked for their feedback on their ability to understand questions, whether anything important to them was missing and if they would be happy providing this type of data via questionnaire in routine MSK care.

Confidentiality: The survey was fully anonymised so that no identifiable information was collected or retained within the Health Survey software.

Right to Withdraw: Respondents had the right to withdraw at any time and if they did so without completing the survey then their results were excluded from the analysis.

CONSENSUS STAGE 3: STAKEHOLDER ENGAGEMENT:

Email invitations were sent to health professionals who had contacted the research team a priori who had expressed interest in the project, and to professional networks with expertise in this area.

Communications: the survey invitation and link was also included in the Versus Arthritis e-bulletin Network News in September 2020 (distribution n=7000), and invites were posted on social media (Twitter/Linked In) by the research team. The survey detail and link were also added to an e-bulletin for the University Research User Group (RUG) to facilitate further patient/service user uptake.

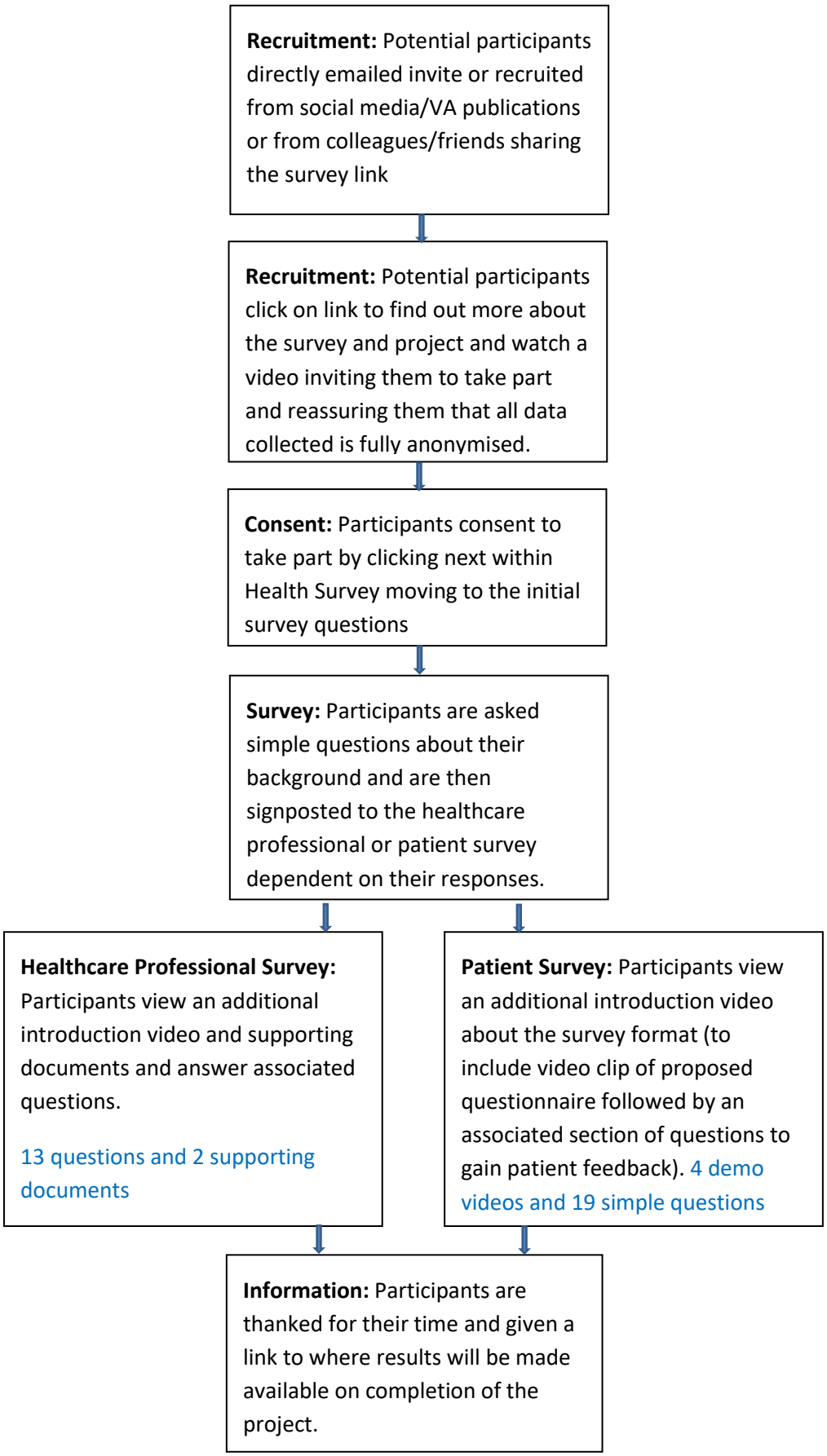
Software settings prevented the completion of the survey multiple times from the same internet location in order to limit the number of responses from the same individual.

See **Figure 5-1** for recruitment/survey flow chart.

CONSENT

Consent was implied through participants clicking on the link to the survey and after viewing the introduction videos continuing to the next page to start the survey questions. Any incomplete surveys were excluded, giving participants the chance to withdraw at any point up to submitting the survey. Patient information leaflets were also provided.

Figure 5-1: Recruitment/Survey Flow



STAGE 4: ANALYSIS PLAN

Data Analysis: questions and responses to the survey were numbered and coded within the Keele Health Survey software, and on completion of data collection the anonymised data was exported to Microsoft Excel for collation and analysis.

Determining scoring system/definition of consensus determined a priori: Frequency counts for each question were analysed. Consensus was defined a priori as agreement from at least 70% of the voting participants/stakeholders as supported by the COS-STAD development group (Kirkham et al, 2017).

Criteria for inclusion/exclusion of outcomes/metrics: Inclusion/exclusion of metrics/outcome tools was decided through analysis of how participants rated metrics (extremely important, very important, moderately important, neutral, slightly important, low importance, not at all important) (Vagias, 2006), and the 70% pre-determined agreement level, alongside supporting questions on feasibility, and on patient feedback and comments on the proposed patient questionnaire. Analysis involved calculating the percentage of health professionals that rated a metric as extremely or very important, this was deemed as strong consensus support if above the 70% pre-defined level and calculating the percentage of respondents who rated the metric as extremely, very or moderately important, if this was above 70% but did not meet the strong consensus category then this was rated as moderate consensus. The strong consensus metrics would then be used to inform the mandatory metrics and the moderate consensus to help inform mandatory and optional metrics supported by feedback from patients and healthcare professionals. The objective of the consensus process was to reduce rather than increase metrics from those listed, but opportunity was given for both healthcare professionals and patients to list anything else important to them that they felt was missing. If a category such as demographics, clinical factors etc had no metric that met strong consensus then the highest scoring metric was considered for inclusion within the mandated set.

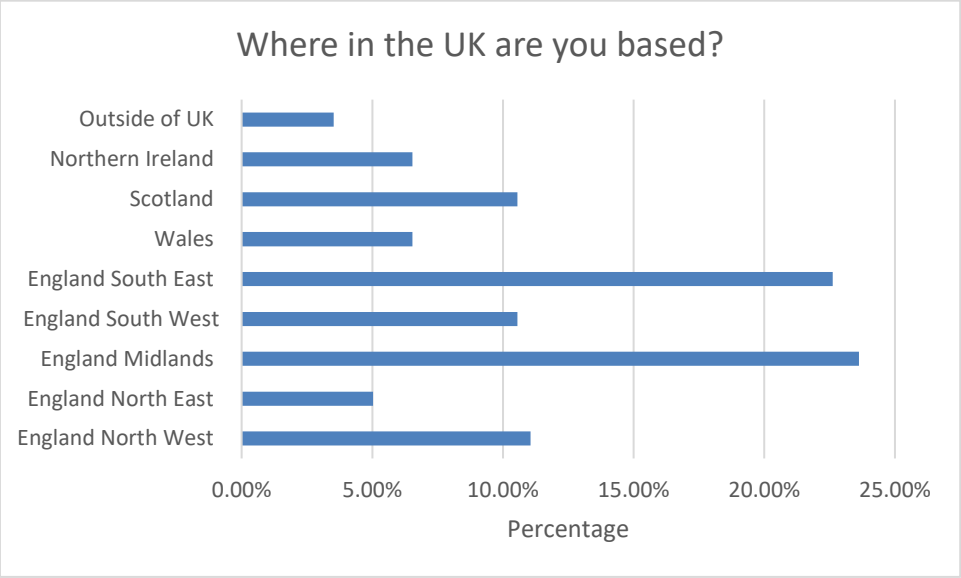
Results

RESPONSE RATES:

There were 199 complete surveys (166 healthcare professionals and 33 patients) collected between September 2020 and January 2021, and 221 incomplete surveys which were automatically excluded from the analysis as this was deemed as not consenting to participate (see consent section above).

There was good spread across the UK with highest uptake from the Midlands (24%) and South East (23%) and lowest uptake from outside the UK (4%), and the North East (5%) (see **Figure 5-2**).

Figure 5-2: Uptake of survey across the UK



Health Professional Survey

The majority of healthcare professional respondents were Physiotherapists and of those most worked in the NHS (**Table 5-4**). Of the 17 that classed themselves as 'other', professional backgrounds included; MSK researcher, Orthopaedic Surgeon, Sports Medicine Specialist, Podiatrists, Director of MSK Services, Osteopath, Massage/Pain therapist and a Podiatric Surgeon.

Table 5-4: Demographics of survey respondents

What is your professional background?		
Answer	Count	Percentage
GP	10	6.02%
FCP (First Contact Physiotherapist)	46	27.71%
Physiotherapist (NHS)	98	59.04%
Physiotherapist (Private Organisation/Practice)	15	9.04%
Physiotherapist (MOD)	6	3.61%
MSK Service Manager	19	11.45%
Other	17	10.24%

Note: respondents could tick more than one answer

Results for healthcare professional consensus on key MSK metrics are shown in **Table 5-5**.

Healthcare professionals were asked to; 'state how important you feel each metric is for inclusion in the standardised MSK dataset.' Variables that met strong or moderate consensus are highlighted within consensus columns.

Table 5-5: Health Professional Consensus Agreement on MSK Metrics

								Consensus	
	Extremely Important	Very Important	Moderately Important	Neutral	Slightly Important	Low Importance	Not at all Important	Strong consensus	Moderate consensus
Demographics									
Age	54.2	29.5	12.1	2.4	1.2	0.6	0	83.7	95.8
Sex	34.3	27.1	24.7	10.2	2.4	1.2	0	61.4	86.1
Education	10.8	22.3	30.1	22.9	5.4	6.6	1.8	33.1	63.2
Ethnicity	20.5	21.7	31.3	15.1	4.2	4.2	3.0	42.2	73.5
Clinical Factors									
Pain site	61.5	27.1	9.0	1.2	1.2	0	0	88.6	97.6
Comorbidities	53.0	38.6	8.4	0	0	0	0	91.6	100
Duration of symptoms	54.8	36.8	6.6	1.8	0	0	0	91.6	98.2
Previous surgery	26.5	36.1	27.1	4.8	4.8	0.6	0	62.6	89.7
Self-reported as disabled	29.5	34.9	20.5	9.6	3.6	1.8	0	64.4	84.9
Previous physiotherapy	20.5	34.9	30.1	6.6	5.4	2.4	0	55.4	85.5
Assisted with questionnaire	14.5	17.5	27.1	24.1	5.4	10.8	0.6	32.0	59.1
Employment									
Work status	38.0	38.6	18.1	3.0	1.8	0.6	0	77.6	94.7
Work absence	37.4	39.2	18.1	3.0	2.4	0	0	77.6	94.7
Work absence duration	30.1	42.8	20.5	1.8	3.6	1.2	0	72.9	93.4
Benefit status	15.1	33.1	33.1	12.1	1.8	3.6	1.2	48.2	81.3
Functional Status/PROMs									
MSK-HQ	30.1	31.9	21.7	11.5	1.8	1.8	1.2	62.0	83.7
NPRS	24.7	27.1	30.1	10.8	2.4	3.0	1.8	51.8	81.9
STarT MSK	20.5	38.6	20.5	12.1	2.4	3.6	2.4	59.1	79.6
EQ-5D-5L	9.6	22.9	26.5	23.5	3.0	6.6	7.8	32.5	59.0
Patient Experience/PREMs									
VPAI Care and Respect	47.0	38.6	6.6	4.2	0.6	1.8	1.2	86.6	92.2
VPAI Understood and Valued	48.8	37.4	7.2	4.2	0	1.8	0.6	86.2	93.4
FFT	31.9	44.0	13.3	7.2	1.2	1.2	1.2	75.9	89.2
Confidence in clinical competence	41.0	32.5	15.1	6.6	1.8	1.8	1.2	73.5	88.6

CollaboRATE/Shared decision making	39.7	40.4	10.8	6.0	1.2	0.6	1.2	80.1	90.9
Given sufficient information	35.5	40.4	12.1	9.0	0.6	1.2	1.2	75.9	88.0
Global improvement	31.9	41.0	16.3	7.8	1.2	0.6	1.2	72.9	89.2
Understanding of health condition	37.4	42.2	13.3	5.4	0	0.6	1.2	79.6	92.9
Confidence to manage yourself	48.8	36.1	10.2	3.0	0	0.6	1.2	84.9	95.1
Timeliness and Convenience	17.5	36.1	31.9	10.2	3.0	0.6	0.6	53.6	85.5
Economic Factors									
Health utilisation	19.9	36.8	30.1	10.2	1.2	1.2	0.6	56.7	86.8
Investigations and treatments	22.9	43.4	24.1	6.0	3.0	0.6	0	66.3	90.4
Inpatient stays	6.6	27.7	40.4	14.5	6.6	3.6	0.6	34.3	74.7
Prescribed medication	27.7	38.6	29.5	3.0	1.2	0	0	66.3	95.8

FFT; Friends and Family Test, MSK-HQ; Musculoskeletal Health Questionnaire, NPRS; Numeric Pain Rating Scale, VPAI; Valuing Patients as Individuals.

Strong consensus = extremely important + very important; Moderate consensus = extremely important + very important + moderately important

FURTHER QUESTIONS FOR HEALTH PROFESSIONALS

Additional metrics: 25% of healthcare professionals were collecting additional metrics for MSK patients. There were 39 comments. These mainly focused on PROMs collected in practice with 9 respondents (5.4%) stating they collected the Patient Specific Functional Scale (PSFS) in practice, 2 reported collecting the Brief Pain Inventory (BPI), and other respondents listed a variety of condition specific tools such as the Oxford Hip Score (OHS) and Oxford Knee Score (OKS), the Knee injury and Osteoarthritis Outcome Score (KOOS), the Hip injury and Osteoarthritis Outcome Score (HOOS), the Oswestry Disability Index (ODI) and Neck Disability Index (NDI). Other generic tools mentioned included the Tampa Scale for Kinesiophobia, Measure Yourself Medical Outcome Profile (MYMOP) and STarT Back alongside collection of red and yellow flags.

Collection of metrics in routine practice**:** It can be seen from the results that there is wide variation in how these metrics are currently being collected in practice (**Table 5-6**). The metrics that are less frequently collected are economic factors and patient experience measures and a significant number of PROMs and PREMs are still being collected on paper.

Table 5-6: Metric collection in routine practice

How are you collecting these metrics? (%)				
	On paper	Through Electronic Patient Record (EPR)	Through electronic platform/app (not EPR)	Not collecting
Demographics	14.5	65.7	12.1	7.8
Clinical Factors	21.7	59.0	11.5	7.8
Employment	17.5	53.6	12.0	16.6
Functional Status/PROMs	25.9	44.6	12.7	16.9
Patient Experience/PREMs	30.7	24.1	19.9	25.3
Economic Factors	13.9	44.0	9.0	33.1

Timing and purpose of metric collection are shown in **Table 5-7** alongside usefulness of standardised data collection and feasibility of collecting these metrics in practice.

Table 5-7: Healthcare Professional views on metric collection

When are you collecting these metrics?*		
Answer	Count	Percentage
At referral	68	40.96%
When appointed	24	14.46%
At first appointment	143	86.14%
At discharge	65	39.16%
At certain time point post treatment	30	18.07%
Not collecting	6	3.61%
Other	6	3.61%
What do you currently use these metrics/data for?		
Answer	Count	Percentage
Monitor patient progress	107	64.46%
Provide patient summary of progress	59	35.54%
Communicate patient progress with clinical teams	62	37.35%
Local service evaluation	104	62.65%
Local audit	98	59.04%
Research projects	30	18.07%
Benchmarking with other services	24	14.46%
As Key Performance Indicators (KPIs) to report to CCG	72	43.37%
To evaluate resource use	30	18.07%
None of the above	11	6.63%
Do you feel that standardised data collection to allow for national benchmarking of MSK community and primary care services would be useful to the MSK community to help drive improvements in care for MSK patients?		
Answer	Count	Percentage
Very Useful	106	63.86%
Useful	45	27.11%
Unsure	11	6.63%
Not Useful	4	2.41%
Not Useful at all	0	0.00%
Would you be able to implement our Keele standardised MSK dataset/core outcome set collection within your service/practice?		
Answer	Count	Percentage
Yes	66	39.76%
No	6	3.61%
Maybe	94	56.63%
Comments	47	28.31%

* Multiple response question

A large amount of MSK healthcare professionals are capturing data at the first appointment for patients (86%) but less at discharge or at a specified time point post treatment (57% in total) and 4% are not collecting at all. Metric collection is largely used for local service evaluation and audit, for monitoring patient progress and to submit as key performance indicators (KPIs) as part of quality reporting to commissioners.

91% of respondents thought that a standardised data collection would be useful to the MSK community to drive improvements in patient care.

40% of respondents thought they could implement the core outcome set developed with a further 57% thinking that they may be able to implement, giving 97% who could potentially take this COS into practice. Comments included; that implementing would take time to setup, it would need to fit with current system setup, it could be implemented if IT systems allowed and metrics could be added to EPR systems/embedded, if digital integrated collection, some had already started using the listed metrics in practice whilst others thought it was very different to what is collected currently in primary care, time was listed as a significant barrier alongside cost of collection and compliance with collection, and organisational support and a national drive to collection were seen as key for successful implementation.

Patient Survey

33 patients completed the patient survey. Of these 2 patients were under 18 and therefore were not able to continue with the full survey and a further 6 had not had a recent experience of accessing or receiving care (in the last 12 months) and therefore were also not able to continue with the full survey. 25 patients answered all of the survey questions.

21 patients (84%) had joint, muscle, or back symptoms/aches/pains, 2 (8%) were post MSK surgery and 2 (8%) had inflammatory conditions showing highest representation for non-surgical non-inflammatory patients which is the main focus of the COS.

36% of patients stated that they had to fill in a questionnaire before and/or after their appointment/treatment. **Table 5-8** shows patient views on the proposed questions that would need to be asked to gather data on their health condition, work, experience and healthcare visits (see **Appendix 5-3** for survey detail). Patient comments are also included in the appendix.

Table 5-8: Patient Views on Proposed Questionnaire

Questions % Yes (count)	Were these questions easy to understand	Was there anything included in this section that you would be unhappy to share	In this section is there anything you feel is missing
Survey Sections			
About your condition	92% (23)	4% (1)	56% (14)
About your work	92% (23)	8% (2)	32% (8)
About your experience	92% (23)	4% (1)	24% (6)
About your healthcare visits	96% (24)	4% (1)	16% (4)

The patient survey showed that the majority of respondents (92-96%) found the proposed questions needed to collect data for the COS to be easy to understand, and there were very few questions included that patients felt they would be unhappy to fill in/share. The only areas that patients were less confident including were benefit and employment status. A number of patients highlighted that they would have liked more questions around anxiety and depression to capture the impact their MSK condition is having on their mental health and work questions may need to be adapted to consider discomfort at work and to take into consideration time off work due to inhibited function rather than purely due to pain.

Completion of Questionnaires: 88% of patients said they would be happy to complete a questionnaire like this when visiting their GP/Physiotherapist about their MSK problem.

Usefulness: 88% of patients (22/25) thought that answering questions like this would help guide their clinical consultation (thinking about face to face, telephone and virtual/online experiences). 76% of patients thought that a standardised MSK dataset would be useful to drive improvements in patient care.

Discussion

Response rates for healthcare professionals exceeded expectations, with uptake widespread across the UK. This supports the interest in and importance of this topic in the current healthcare

climate and the need for clear guidance and recommendations to support development of consistent and continuous data capture within the clinical setting.

The metrics that reached a strong level of consensus support for inclusion within a standardised MSK dataset were; **age**, **pain site**, **comorbidities**, **duration of symptoms**, **work status**, **work absence**, **work absence duration**, PREMS; **VPAI** (care and respect and understood and valued), **FFT**, **confidence in clinical competence**, **CollaboRATE** (shared decision making), **given sufficient information**, **global improvement**, **understanding health condition** (MSK-HQ Q12), and **confidence to manage yourself** (MSK-HQ Q13). No PROMs or economic factors met strong consensus.

The additional metrics which reached moderate consensus were; **sex**, **ethnicity**, **previous surgery**, **self-reported as disabled**, **previous physiotherapy**, **benefit status**, **MSK-HQ**, **NPRS**, **STarT MSK**, patient experience domain of **timeliness and convenience**, and economic factors of **health utilisation**, **investigations and treatments**, **inpatient stays** and **prescription medications**.

Metrics which did not meet strong or moderate consensus were; education, assisted with questionnaire and the EQ5D5L. Education and assistance with questionnaire are both potential variables for use in case-mix adjustment models (Burgess et al, 2019, 2020a, **Chapters 2-3**), but are not feasible for inclusion in modelling if there is minimal support to collect them in routine practice. The EQ5D5L is a widely used health status measure for clinical and economic appraisal (EuroQol, 2021). Price et al, (2019) showed that it is not as responsive as the MSK-HQ in MSK patients. Unlike other tools however it allows for the calculation of QALYs (Quality Adjusted Life Years) for use in health economic analyses (Euroqol, 2021) which is a limitation to exclusion.

In the initial evidence informed list of metrics both the MSK-HQ and pain intensity (NPRS) were part of the mandated recommendation. None of the included functional status metrics/PROMs reached strong consensus. MSK-HQ however achieved the highest consensus, reaching 63%

agreement (extremely important/very important) and 84% agreement at the moderate level (extremely important/very important/moderately important) and had strong consensus for inclusion within patient experience metrics (Q12 and Q13). It is essential that a standardised PROM is part of the dataset and based on these results the inclusion of the MSK-HQ is recommended. The PSFS was also listed by 5% of healthcare professionals as part of their current MSK metric collection. This tool is similar to the MSK-HQ in the way that it is a generic PROM applicable to a number of conditions, but it also allows for patients to identify specific activities that they are finding difficult due to their condition (Stratford et al, 1995). The addition of this PROM as an optional metric may therefore be useful.

Collection of MSK metrics currently is highly varied, with different metrics collected at different time points in different ways. Without standardisation of which metrics are collected and at which time-points in routine care, making comparisons in patient outcome with regards to quality and benchmarking is difficult. Baseline patient factors are necessary for case-mix adjustment in order to adjust for varying complexity across providers and allow for fair comparison. It is important that factors used for this purpose are simple to extract from electronic medical records or to collect from patients for this to be feasibly undertaken in routine practice. From previous evidence the most predictive patient factors of MSK functional outcome widely used in case-mix adjustment are; baseline PROM score, comorbidities, age, disability, duration of symptoms, employment/work absence, surgical history, and socio-economic status (Burgess et al, 2019, 2020a, **Chapter 2-3**). Of these, age, comorbidities, duration of symptoms, and employment/work absence reached strong consensus for inclusion within an MSK COS. Baseline PROM score would also need to be included for the outcome adjusted within a case-mix model.

Seventeen percent of respondents reported not collecting any PROMs and 26% not collecting any PREMs which shows that there is significant work needed to support the use of PROMs and

PREMs in routine MSK care and to develop the infrastructure to consistently collect an MSK COS as part of routine practice. Barriers to current and future collection included IT infrastructure and integration of digital systems, time, resources, and support from managers/organisation/stakeholders. Comments from both healthcare professionals and patients included needing to clearly demonstrate the use of the data as too often questionnaires are collected but not used by clinicians to support clinical decision making or fed back to staff or stakeholders with regards to the quality of care delivered. It is therefore clear that stakeholders would want to see the value realised with collecting this core set of data with demonstrable improvements in patient care.

The collection of PREMs was clearly seen as a high priority by clinicians reflected in the fact that 9 out of 10 PREM domains met strong consensus for inclusion making up a significant proportion of those metrics with strong consensus. Interestingly clinicians did not see 'timeliness and convenience' as such an important domain, this may be because they feel this area is beyond their control as clinicians and is a metric more focused to system setup. Patients however felt that another system focused metric around access to care was an important metric to consider for inclusion.

Other general comments included the difficulty currently in accessing face to face physiotherapy care due to the COVID-19 pandemic and how difficult it is to ask questions to clinicians following reflection on appointments due to the pressure from clinicians to discharge. Digital platforms and apps offer an opportunity to improve communication channels between patients and clinicians and could offer significant support with current pressures on face to face delivery of care, and as services adapt and develop in the future with further digital innovation and integration.

STRENGTHS AND LIMITATIONS

Strengths: The strengths of this study were that the aims to develop consensus, the method of a single round survey, participant engagement, and the consensus definition were all determined a priori meeting the COS-STAD and Delphi quality criteria/recommendations (Kirkham et al, 2017, Diamond et al, 2014). Failure to adequately define and use clear criteria for consensus can challenge the notion of consensus itself (Diamond et al, 2014).

Limitations: The survey was limited to the English language and to digitally literate respondents. A large number of incomplete questionnaires (221) were excluded from the analysis. Inviting and encouraging respondents to view online surveys is relatively easy but the capture of complete data is far more challenging (McRobert et al, 2018) as was demonstrated within this study. Another potential limitation is the use of a single round online survey rather than a conventional multiple round Delphi survey where initial consensus findings are then presented back to respondents for further ratification. The COMET guidelines discuss the variation in approaches to COS development and have identified that further methodological guidance on the optimal approach would be helpful (Williamson et al, 2017). Results for this consensus study will need to be discussed with national policy makers to identify if any variables that met moderate but not strong consensus such as ethnicity are deemed essential as part of national reporting and therefore would need to be included within the final recommendation.

Conclusion

Seven baseline patient factors reached strong consensus for inclusion into an MSK COS for use in community and primary care MSK services alongside an MSK PROM. From the generic PROMs included the MSK-HQ reached the highest level of support and is therefore recommended for inclusion. Nine out of ten patient experience measures had strong consensus agreement showing the importance clinicians place on capturing information on all aspects of patient care. Interestingly collection of PREMs although strongly supported for collection using a number of metrics was the main area that was not consistently being collected in current practice.

Moderate consensus support was also reached for collection of a significant number of additional factors in each category. Items that could be removed from the list following the consensus process include; education, assistance with questionnaire, and the EQ-5D. Further consideration needs to be given to the wording around work metrics and to the inclusion of a mental health metric following patient feedback.

There was overwhelming support from healthcare professionals and patients that standardising data collection in this way would improve patient care. Further work is needed to set out the specific detail of the agreed COS which is presented in a separate Chapter/paper (**Chapter 7**), supported by; testing of a feasible case-mix model (**Chapter 6**), how and when to collect agreed metrics including specific wording of questions for patients, and how to utilise this data effectively for service evaluation, audit, benchmarking, and quality improvement.

6 Chapter 6: Musculoskeletal case-mix adjustment in a UK primary/community care cohort: Testing musculoskeletal models to make recommendations in this setting.

Chapter 6 took findings from **Chapter 2** and **Chapter 3** and looked to test both existing case-mix models identified in **Chapter 2**, and explore inclusion of additional case-mix variables identified within **Chapter 3** within a secondary analysis of a newly collected dataset collected as part of a randomised controlled trial (RCT) within the Research Institute.

Abstract

Benchmarking musculoskeletal (MSK) services is limited by the need to adjust for differences in patient characteristics/case-mix. Without this providers and services cannot be usefully compared. This chapter investigates the predictive ability of case-mix adjustment models in a primary/community care cohort and makes recommendations for future case-mix adjustment and benchmarking in this setting.

Objectives: To investigate the predictive ability of two existing MSK case-mix adjustment models and compare to the predictive ability of an evidence-informed and statistically-informed model.

Method: A secondary analysis of the STarT MSK (Subgrouping for Targeted Treatment in MuSculosKeletal conditions) cluster randomised controlled trial data (n=1211). Stepwise linear regression models were built and compared using available baseline variables. The MSK-HQ was used as the primary functional status outcome.

Results: Two existing models were compared (UK National PROMs Model and US FOTO Model) using available variables. Of these two models the modified US FOTO model showed the best predictive ability in this cohort predicting 44% of the variation in MSK-HQ outcome, the modified UK National PROMs model predicted 41%. A newly developed evidence-informed model (Keele Model 1) performed no better than the existing models, and a statistically-informed model (Keele

Model 2) gave only an additional 2% increase in model power compared to the modified US FOTO model.

Conclusion: All models showed strong predictive ability. The modified US FOTO model looks to be best suited to the UK primary/community care cohort of the existing models. This model performed so well that we recommend that this adapted FOTO model is used in a UK setting moving forwards rather than development of an alternative UK model.

Contribution to the Thesis:

- This chapter validates two modified MSK case-mix adjustment models in primary/community care.
- It shows that newly developed evidence-informed and statistically-informed models also performed well in this community/primary care cohort.
- An agreed case-mix adjustment method is needed alongside standardised outcome collection (discussed in **Chapter 5**).
- Implementation of case-mix adjustment methods would allow for national benchmarking.
- National benchmarking in community MSK would help to drive quality improvement/innovation.

The contents of this chapter have been published in part in:

Burgess, R., Lewis, M., and Hill, J.C., 2021. Musculoskeletal case-mix adjustment in a UK primary/community care cohort: Testing Musculoskeletal models to make recommendations in this setting. *Musculoskeletal Science and Practice*, https://authors.elsevier.com/sd/article/S2468-7812(21)00139-9

This paper has also been presented as a poster at PUK 2021.

Introduction

Case-mix adjustment is a statistical process that aims to account for differences in the mix of patient attributes across definitive patient cohorts (e.g., patient groups treated by different healthcare providers), in order to make fair comparisons of the relative effectiveness (outcome) of care provided (Iezzoni, 2009). This adjustment for case-mix is important as there is strong evidence demonstrating that patient factors such as worse baseline function, higher symptom/pain severity, worse mental wellbeing, more comorbidities, and older age can have a detrimental effect on MSK treatment outcome (Burgess et al, 2020, **Chapter 3**). If known patient attributes that affect treatment outcome are not taken into consideration and 'adjusted out' using statistical modelling, then comparisons of provider outcomes will be biased in favour of those treating less complex patient groups and would not allow for effective or fair benchmarking. To be genuinely comparable the data needs to be adjusted for the mix of cases that each individual provider has (Nuttall et al, 2015).

Benchmarking in clinical practice represents a process by which individual providers can compare and share best practice, facilitating continuous quality improvement (Siemens et al, 2017). Routine data collection within the UK National Patient Reported Outcome Measures (PROMs) Programme has demonstrated the value of benchmarking using case-mix adjustment to identify providers and specific treatment approaches that are delivering changes in clinical outcomes that are not typical, known as 'outliers'. For example, arthroplasty implant brand, early post-operative mobilisation regimes, and wound management protocols were seen to impact on clinical outcomes. Significant improvements in widespread clinical outcomes were then realised with the implementation of associated clinical changes to align with positive outliers (Baker et al, 2012, NHS England, 2016, NHS Digital, 2018). Similar comparative data analysis approaches for community and primary care MSK services however are not available.

A previous systematic review (Burgess et al, 2019, **Chapter 2**) identified two existing and distinct MSK case-mix adjustment models. One model was developed by Coles et al (2010) on behalf of the UK Department of Health (DoH) (now NHS England (NHSE)) and is called the National PROMs (NPROMS) model within this chapter, and one was developed by Hart and Connelly (2006) and has been continuously re-validated and improved by the Focus on Therapeutic Outcomes (FOTO) US research team (Deutscher et al, 2018). The UK NPROMS model was developed and continues to be used to adjust MSK surgical outcomes including hip and knee joint replacement surgery in the UK. The FOTO model was developed and continues to be used to case-mix adjust outcomes for patients referred to MSK rehabilitation outpatient clinics and is now used by more than 23,000 clinicians across more than 12,000 clinics throughout all 50 US states (FOTO, 2021). There is however no existing or validated MSK case-mix model for use in UK primary/community healthcare.

Stepwise regression models remain controversial as they allow priority to statistical criteria for inclusion into a model rather than basing model development on theoretical research criteria (Bryman and Cromer, 2001) and can therefore give rise to questionable external validity. The standards for reporting of statistical models outlined by Krumholz et al (2006) recommend that case-mix adjustment models should be informed by clinical judgement and insights from published literature with regards to the selection of candidate variables. This should allow for the development of coherent case-mix models and should minimise the idiosyncrasies of individual datasets (Krumholz et al, 2006). These concepts and recommendations are explored and discussed within the developmental models within this chapter. An additional challenge with case-mix adjustment model development and benchmarking is whether to account for 'nesting' or 'clustering' of patients within clinics (Deutscher et al, 2018). Using a mixed/multi-level model that includes both fixed effects (FE) and random effects (RE) with clinic or provider treated as a RE will explain more variance and therefore improve model prediction (Yen et al, 2015). Use of multi-level models however could adjust out differences in patient outcome attributable to

patient care, and therefore mask differences in quality between providers that are the target of ranking/benchmarking (Deutscher et al, 2018).

This Chapter sought to explore an evidence synthesis approach and a statistical approach alongside testing existing MSK case-mix adjustment models in order to make recommendations and inform case-mix adjustment modelling within a UK primary/community healthcare setting. Mixed models were also explored for existing models as part of the sensitivity analyses.

The specific objectives were to:

1; Explore the predictive ability/validity of a modified NPROMS and a modified FOTO case-mix adjustment model applied to a UK primary/community care patient cohort (modified due to availability of variables and slight differences in how these variables are collected) (Coles, 2010, DoH, 2012, Deutscher et al, 2018),

2; Develop a new case-mix adjustment model (Keele Model 1) using an evidence-synthesis approach (informed by the systematic review detailed in **Chapter 2** (Burgess et al, 2019) and the umbrella review detailed in **Chapter 3** (Burgess et al, 2020), identifying case-mix variables/predictors of functional outcome respectively).

3; Develop a new case-mix adjustment model (Keele Model 2) using a stepwise statistical approach (using all available variables within the dataset) to identify the most parsimonious model within this community/primary care cohort.

Method

Methodological Quality: Methods for this study followed recommendations detailed in 'The Standards for Statistical Models Used for Public Reporting of Health Outcomes' reported by Krumholz et al (2006). The reporting of regression analyses also followed standards detailed by SAMPL Guidelines (Lang and Altman, 2015).

A secondary analysis of prospectively collected data from adult (>18) patients presenting in primary care with MSK pain (back, neck, shoulder, knee, widespread pain) was conducted. This data was collected within the STarT MSK (Subgrouping for Targeted Treatment in MuSculosKeletal conditions), cluster randomised controlled trial (RCT) in 2019/2020 (ISRCTN15366334 (Hill et al, 2020)). A standardised set of metrics were collected for included patients (Hill et al, 2020), these included patient characteristics/demographics, baseline clinical factors, patient reported outcome measures (PROMs), and employment factors. The Musculoskeletal Health Questionnaire (MSK-HQ) functional status PROM was collected on presentation to primary care and again at 6-month follow up. The MSK-HQ has been shown to be valid, reliable, and responsive as a measure of MSK health status in a UK community/primary care setting (Hill et al, 2016, Price et al, 2019, Scott et al, 2020). The STarT MSK Trial data included the following patient factors for evaluation for inclusion within case-mix adjustment models; MSK health/functional status at baseline presentation (MSK-HQ PROM) (continuous 0-56, low to high functioning), age (continuous), sex (male/female), ethnicity (6 categories), socioeconomic status (SES) (health literacy (5 categories), Index of Multiple Deprivation (IMD 10)), symptom duration (whole month without pain (7 categories)), pain site (5 categories), pain intensity (continuous 0-10), distress (continuous 0-10), self-efficacy (continuous 0-10), previous pain episodes (5 categories), previous surgery (yes/no), living alone (yes/no), paid employment (yes/no), work absence (yes/no), work absence duration (continuous), comorbidities (list of 12 (4 categories)), physical activity (8 categories), disability (EQ5D5L PROM), fear avoidance beliefs (FAB) (FAB-TSK PROM), see **Table 6-1** for detail of variables across models (see STarT MSK Trial protocol paper (Hill et al, 2020) for further detail of RCT). SPSS software was used for ordinary least squares (OLS) regression analysis.

Multiple linear regression equation with k predictor variables: $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots \beta_k X_k + E$

Table 6-1 Variables in STarT MSK Trial Dataset & those used in NPROMS and FOTO Models.

Baseline Variables	**Available in STarT MSK Trial Dataset**	**NPROMS Model (DoH/NHSE, 2012/2013)**	**FOTO Model (Deutscher 2018)**
Characteristics			
Age	X	X	X
Sex	X	X	X
Ethnicity	X	X	
SES-IMD	X	X	
SES-Education			
SES-Job Title	X		
SES-Health Literacy	X		
Baseline Clinical Factors			
Pain Site/Body Part/Impairment	X	X	X
Comorbidities	X	X	X
Duration of Symptoms	X	X	X
Previous Surgery	X	X (2012)	X
Previous Pain Episode	X		
Previous Treatment			X
General Health		X (2012)	
Medication at Intake			X
Employment			
Work Status	X		
Work Absence	X		
Work Absence Duration	X		
Payer			X
Function/Disability			
Pain Intensity	X		
Functional PROM Score (e.g. MSK-HQ)	X	X	X
Disability/QoL PROM (e.g. EQ5D)	X	X	
Self-reported as Disabled		X	
Psychological Distress	X		
Fear Avoidance (FABQ, TLS-11)	X		
Self-Efficacy	X		
Physical Activity	X		X
Social Factors			
Living Alone	X	X	
Support Needed with Questionnaire		X	
Model Power			
R^2 (%)	NA	23-30%	30-40%

Modified FOTO Model

The FOTO model we aimed to emulate is described by Deutscher et al (2018). Available variables in the STarT MSK Trial dataset included: age (continuous), sex (binary), baseline functional status (FS) (MSK-HQ score (continuous) replaced FOTO FS computerised adaptive testing (CAT) PROM), pain site (5 original categories dichotomised to single or multisite pain (replaced body part and care type i.e., orthopaedic)), duration of symptoms (7 categories (how long since whole month without pain) replaced 6 acuity categories categorized as number of days from onset of the treated condition), comorbidities (4 categories), previous pain episodes (5 categories (replaced previous treatment)), physical activity (8 categories), employment (current paid employment (binary) replaced payer as not relevant for UK NHS data), previous surgery (binary due to low numbers with multiple surgeries). The only variable missing that is included within the FOTO model described by Deutscher et al (2018) is medication at intake. This variable however was not a key variable within the model (Deutscher et al, 2018). The most important predictors in the model developed by Deutscher et al (2018) were intake FS, symptom duration, payer, and age.

It is important to note that specific categories are different for some variables therefore this is not an exact replica of Deutscher et al's model (2018). Timescales for collection within the STarT MSK trial were also standardised at baseline and 6 month follow up which is not the case in the FOTO data collection where data is collected at discharge.

Case-mix adjustment: A backwards stepwise ordinary least squares (OLS) regression model was used to emulate the method used by Deutscher et al (2018). Variables were entered into the model if $p<0.05$ and removed from the model if $p>0.1$ (these values differ from those used by Deutscher et al (2018) due to the significantly smaller sample size (entered $p<0.005$, removed $p>0.01$)).

Model predictive power was then calculated and displayed showing R^2 and adjusted R^2 values and ANOVA's calculated to provide F values to demonstrate model fit. The model power was also compared with a model using an alternative PROM (EQ5D5L).This alternative PROM was used to additionally evaluate both of the existing models (modified FOTO and modified NPROMS) due to the MSK-HQ (primary PROM for use in UK community/primary care MSK) being different to those used in these internally validated models.

Modified NPROMS Model

Variables for the NPROMS method were taken from the DoH (2012) and NHS England (2013) publications on case-mix adjustment methodology. We aimed to emulate the variables in the most recently published model (NHS England, 2013) but also included surgical history as this was still relevant to a community/primary care dataset unlike the 2013 NPROMS dataset where surgical revisions were separated out from the main dataset (NHS England, 2013).

Available variables in the STarT MSK Trial dataset included: age (continuous), sex (binary), ethnicity (6 categories), baseline FS (MSK-HQ score (continuous) used instead of EQ5D and Oxford Hip and Knee Scores), Index of Multiple Deprivation (IMD 10), pain site (5 original categories dichotomised to single or multisite pain (replaced impairment)), duration of symptoms (7 categories (how long since whole month without pain)), comorbidities (4 categories), previous surgery (binary), living alone (binary). The variable of self-reported as disabled and support needed with questionnaire were missing from the STarT MSK Trial dataset but the EQ5D5L index score was available as a measure of disability/quality of life so was added to the included variables. Ethnicity although available was not entered into the model due to minimal variation across categories (97% white), this follows the method reported by Coles in the original NPROMs development paper (Cole, 2010).

It is important to note that specific categories are different for some variables therefore this is not an exact replica of the NPROMs methods (Coles 2010, DoH 2012, NHS England 2013).

Case-mix adjustment: In order to compare with the previous modified FOTO model, we used an OLS stepwise model as used in the DoH (2012) NPROMS method. Variables were entered into the model if $p<0.05$ and removed from the model if $p>0.1$.

Model predictive power was then calculated and displayed showing changes in R^2 and F values. The modified model power was also compared with a model using an alternative PROM (EQ5D5L) to evaluate if R^2 values were comparable across PROM measures.

Keele Model 1: Evidence informed model

This Keele Model 1 was developed using theoretical criteria identified in previous evidence syntheses (Burgess et al, 2019, 2020, **Chapter 2-3**).

Independent variables were force entered in batches based on evidence-based recommendations (see **Table 6-2** for identified case-mix variables and strength of evidence). Of these variables all of the 'highly recommended' variables were available in the dataset and 4 out of 7 of the moderate evidence 'recommended' variables were available and initially entered into the model. Variables with very strong evidence made up 'model version a', variables with strong evidence were added to make 'model version b' and variables with moderate evidence were added (where available) to make 'model version c'. These were added in SPSS using the 'next' function in linear regression to allow for independent variables to be split into groups and entered into the model in group (hierarchical) order in an evidence-informed rather than in a statistically informed way.

Table 6-2 Recommendations for variables to include in MSK case-mix adjustment model development

Very strong evidence **Highly Recommended**	**Strong evidence** **Highly Recommended**	**Moderate evidence** **Recommended if feasible to collect in addition**	**Limited evidence** **Not Recommended for initial inclusion**
Age (continuous, 0-120)	Disability (binary self-rated or questionnaire)	Assistance with questionnaire (binary yes/no)	Exercise history (potential list of 3)
Baseline PROM score (continuous)	Duration of symptoms (potential list of 10)	BMI (4 categories)	Fear avoidance (FABQ)
Comorbidities (potential list of 11)	Payer/Employment /Sick leave duration (potential list of 10 or sick leave duration list (to be determined))	Depression/Mental wellbeing (binary/mental wellbeing questionnaire)	General health (potential list of 5)
	Socioeconomic status (IMD 2010 or Education)	Ethnicity (potential list of 5)	Living alone (binary yes/no)
	Surgical history (binary yes/no)	Gender (binary male/female)	Use of medication at intake for condition (binary yes/no)
		Impairment type/anatomical body part (potential list of 10)	Vitality (SF36)
		Pain intensity (NPRS 0-10)	

PROM; Patient Reported Outcome Measure, IMD; Index of Multiple Deprivation, BMI; Body Mass Index, NPRS; Numeric Pain Rating Scale, FABQ, Fear Avoidance Beliefs Questionnaire.

* Combined findings from Burgess et al 2019, 2020.

Keele Model 2: Statistically informed model

The full list of available variables within the STarT MSK Trial dataset listed above (see Table **6-1**) were added to a regression model using a stepwise OLS model approach (20 variables in total were available). Variables were entered into the model if $p<0.05$ and removed from the model if $p>0.1$. In a stepwise model the order of inclusion is determined by the contribution of each variable to the explained variance (Bryman and Cromer, 2001). This model therefore used all available variables and allowed the model to be determined purely by the statistical contribution

of independent variables to the model. A forwards stepwise model was used to display changes in model power with addition of further variables giving associated changes in R^2 and F values and to show what a 'parsimonious' model could look like using as few variables as possible.

Model Assumptions

For all objectives/models detailed above, assumptions of independence were assessed by using Pearson's Correlation Coefficients and collinearity/multicollinearity by assessing Tolerance and Variance Inflation Factor (VIF). Criteria for determining collinearity: correlation>0.8, tolerance statistic<0.1, and VIF>10 (Senaviratna and Cooray, 2019). Normality and homoscedasticity were tested by plotting a normal distribution line against the distribution of residuals and by fitting a regression line to the squared residuals across the predicted outcome (MSK-HQ score at 6 month) respectively (Deutscher et al, 2018).

Sensitivity Analysis

To take account of potential clustering of patients between GP practices (i.e., potential clustering of similar patients within practices leading to larger variability of patient outcomes between practices than within practices), regression analyses were repeated for modified FOTO and NPROMS models using mixed/multi-level models including both fixed and random effects. STATA statistical software was used to run additional multi-level models (with GP practice added as a random effect) to identify if this impacted on coefficient values/significance.

Results:

Descriptive data for the STarT MSK Trial dataset is presented in **Table 6-3** providing mean or percentage values for each variable, standard deviation (SD), and number of participants (n) alongside % of participants with available data for each variable. **Table 6-4** presents univariate predictive values for each available variable within the dataset including standardised coefficients (beta) and p values with p<0.05 indicating significance. **Table 6-4** shows that all variables were

significant in predicting MSK-HQ outcome, except for age and ethnicity, with baseline MSK-HQ score, EQ5D5L index score, distress, work absence duration, previous pain episodes and pain intensity being the most independently predictive variables respectively. The data show that there was very low ethnic diversity within this dataset with 97% of participants being classed as 'white' (**Table 6-3**), this may explain why this variable was not predictive in this cohort.

Table 6-3 Descriptive Statistics

Descriptive Statistics				
	Mean/frequency (%)	**SD**	**N**	**% complete data**
N			1211	
MSK-HQ score baseline (mean)	29.17	10.162	1208	99.75
MSK-HQ score 6m FU (mean)	38.17	11.68	972	80.26
Age (mean)	60.03	15.28	1211	100
Sex (female) (f(%))	714(59)		1211	100
Ethnicity (%) Mixed Asian Black White Other Prefer not to say	 8(0.7) 5(0.4) 4(0.3) 1172(97.3) 7(0.6) 9(0.7)		1205	99.50
SES: Health Literacy (help) (f(%)) Never Rarely Sometimes Often Always	 973(81.7) 102(8.6) 73(6.1) 32(2.7) 11(0.9)		1191	98.35
SES: IMD Decile (f(%)) 1 2 3 4 5 6 7 8 9 10	 19(1.6) 31(2.6) 84(6.9) 109(9.0) 114(9.4) 159(13.1) 171(14.1) 212(17.5) 160(13.2) 153(12.6)		1211	100
Duration (without pain) (f(%)) <3m 3-6m 7-12m 1-2y 3-5y 6-10y over 10y	 306(25.5) 207(17.2) 151(12.6) 147(12.2) 162(13.5) 83(6.9) 146(12.1)		1202	99.26

Previous Pain Episodes (f(%)) 0 1 2-3 4-9 10+	 260(21.6) 144(12.0) 231(19.2) 185(15.4) 385(32)		1205	99.50
Pain Intensity (NPRS) (mean)	6.35	2.217	1208	99.75
Comorbidity Count (f(%)) 0 1 2 3 or more	 386(31.9) 439(36.3) 250(20.6) 136(11.2)		1211	100
Pain Site (f(%)) Knee Neck Back Shoulder Multisite	 379(31.3) 130(10.7) 457(37.7) 130(10.7) 116(9.6)		1211	100
Physical Activity (days per week) (f(%)) 0 1 2 3 4 5 6 7	 299(24.8) 150(12.4) 177(14.7) 167(13.8) 105(8.7) 134(11.1) 42(3.5) 134(11.1)		1208	99.75
Previous Surgery (f(%)) Yes No	 149(12.2) 1019(87.8)		1168	96.45
Living Alone (f(%)) Yes No	 212(17.6) 991(82.4)		1203	99,34
EQ5D (mean index score)	0.557	0.235	1171	96.70
FAB-TSK (mean)	25.07	6.465	1186	97.94
Paid Employment (f(%)) Yes No	 561(48.2) 604(51.8)		1165	96.20
Work Affected (mean) (0-10)	4.66	3.191	649	53.59
Time off work (f(%)) Yes No	 203(31.8) 436(68.2)		639	52.77
Distress (mean) (0-10)	5.82	2.579	1208	99.75
Self-efficacy (mean) (0-10)	5.22	2.548	1210	99.92

f; frequency, NPRS; Numeric Pain Rating Scale, SD; standard deviation

Table 6-4 Univariate Analysis

Baseline Variables		95% Confidence Interval for Beta			
	Unstandardised Coefficient Beta	**Lower bound**	**Upper bound**	**Standardised Coefficient Beta**	**P-value**
Characteristics					
Age	-0.024	-0.077	0.030	-0.028	0.382
Sex [Male]					
Female	-2.099	-3.586	-0.612	-0.089	**0.006**
Ethnicity [Non-white]					
White	0.787	-4.537	6.111	0.009	0.772
SES (Health Literacy) [LT]*	-5.066	-6.006	-4.127	-0.324	**<0.0001**
SES (Health Literacy) [Never]†					**<0.0001**
Rarely	-5.286	-7.833	-2.74	-0.125	**<0.0001**
Sometimes	-9.609	-12.56	-6.66	-0.197	**<0.0001**
Often	-15.54	-20.72	-10.36	-0.181	**<0.0001**
Always	-20.95	-28.22	-13.67	-0.173	**<0.0001**
SES (IMD_Decile) [LT]*	0.744	0.420	1.068	0.144	**<0.0001**
SES (IMD_Decile) [1]†					
2	2.879	-5.280	11.04	0.037	**0.489**
3	1.524	-5.675	8.724	0.031	**0.678**
4	4.150	-2.888	11.19	0.098	**0.248**
5	5.585	-1.421	12.59	0.136	**0.118**
6	3.711	-3.140	10.56	0.110	**0.288**
7	5.685	-1.146	12.52	0.173	**0.103**
8	6.754	-0.035	13.54	0.221	**0.051**
9	5.975	-0.882	12.83	0.175	**0.088**
10	8.572	1.701	15.44	0.247	**0.015**
Clinical Factors					
Pain Site [multisite]†					
Knee	8.057	5.335	10.78	0.322	**<0.0001**
Neck	8.118	4.861	11.37	0.216	**<0.0001**
Back	6.458	3.781	9.135	0.268	**<0.0001**
Shoulder	8.113	4.856	11.37	0.216	**<0.0001**
Pain Site [multisite]†					
Single site	7.412	4.892	9.933	0.182	**<0.0001**
Comorbidities [LT]*	-3.537	-4.252	-2.823	-0.298	**<0.0001**
Comorbidities [0]†					
1	-1.063	-2.778	0.652	-0.044	**0.224**
2	-6.129	-8.099	-4.159	-0.215	**<0.0001**
3	-10.84	-13.24	-8.427	-0.298	**<0.0001**
Duration [LT]*	-1.442	-1.784	-1.100	-0.257	**<0.0001**
Duration [<3m]†					
3-6m	0.866	-1.354	3.085	0.028	**0.444**
7-12m	-0.474	-2.982	2.034	-0.013	**0.711**
1-2y	-4.052	-6.468	-1.636	-0.116	**0.001**

3-5y	-5.328	-7.777	-2.879	-0.151	**<0.0001**
6-10y	-8.331	-11.24	-5.426	-0.191	**<0.0001**
over 10y	-6.888	-9.345	-4.432	-0.194	**<0.0001**
Previous Surgery [no]†	-7.138	-9.307	-4.970	-0.206	**<0.0001**
Yes					
Previous Pain Episodes [LT]*	-2.795	-3.237	-2.353	-0.370	**<0.0001**
Previous Pain Episodes [0]†					
1	-3.437	-5.895	-0.979	-0.096	**0.006**
2-3	-4.896	-7.035	-2.757	-0.165	**<0.0001**
4-9	-7.013	-9.337	-4.690	-0.211	**<0.0001**
10+	-11.55	-13.43	-9.659	-0.465	**<0.0001**
Employment					
Work Status [not working]†					
Paid work	3.351	1.860	4.842	0.143	**<0.0001**
Work Absence [no]†					
Yes	-3.485	-5.686	-1.285	-0.141	**0.002**
Work Absence Duration	-0.168	-0.236	-0.099	-0.391	**<0.0001**
Function/Disability					
Pain Intensity	-1.913	-2.219	-1.608	-0.367	**<0.0001**
MSK-HQ	0.665	0.605	0.724	0.577	**<0.0001**
EQ-5D	27.421	24.713	30.129	0.544	**<0.0001**
FAB-TSK	-0.630	-0.738	-0.523	-0.350	**<0.0001**
Distress	-1.789	-2.046	-1.532	-0.402	**<0.0001**
Self-efficacy	0.933	0.648	1.218	0.202	**<0.0001**
Physical Activity	0.325	0.011	0.640	0.065	**0.043**
Social Factors					
Living Alone [no]†					
Yes	-4.15	-6.061	-2.239	-0.136	**<0.0001**
Dependent variable; 6-month MSK-HQ score					

*Linear trend [LT] taken forward for ordinal data (since the Bayesian Information Criteria (BIC) goodness-of-fit was lower for the numerical model than for the full categorical model)

† Reference category (for categorical variables)

Objective 1:

Modified FOTO Model

The STarT MSK Trial dataset included 1211 patients in total. Of those 1211, 905 (75%) had complete data for available FOTO variables. The model summary is shown in **Table 6-5** alongside the standardised coefficients for the final backward stepwise model (model version d) with all redundant /non-significant variables removed.

Table 6-5 Modified FOTO Backward Stepwise Model: Model summary & coefficients

Model Summary[e] (N=905)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
a.	.666[a]	0.444	0.438	8.709	0.444	71.314	10	894	0.000	
b.	.666[b]	0.444	0.438	8.704	0.000	0.031	1	894	0.861	
c.	.666[c]	0.444	0.439	8.700	0.000	0.042	1	895	0.839	
d.	.665[d]	0.442	0.438	8.706	-0.001	2.347	1	896	0.126	1.953

a. Predictors: (Constant), Age, Sex, Duration of symptoms, Physical activities, Previous surgery, Pain site, Comorbidities, Baseline MSK-HQ score, Previous pain episodes, Baseline: Current paid employment

b. Predictors: (Constant), Sex, Duration of symptoms, Physical activities, Previous surgery, Pain site, Comorbidities, Baseline MSK-HQ score, Previous pain episodes, Current paid employment

c. Predictors: (Constant), Duration of symptoms, Physical activities, Previous surgery, Pain site, Comorbidities, Baseline MSK-HQ score, Previous pain episodes, Current paid employment

d. Predictors: (Constant), Duration of symptoms, Physical activities, Previous surgery, Comorbidities, Baseline MSK-HQ score, Previous pain episodes, Current paid employment

e. Dependent Variable: 6 Months: MSK-HQ score [0 - 56 scale: 0=Worst health; 56=Best health]

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
d	(Constant)	30.094	1.401		21.481	0.000	27.344	32.843
	Current paid employment	2.168	0.605	0.093	3.584	0.000	0.981	3.355
	Physical activities	-0.358	0.128	-0.072	-2.805	0.005	-0.608	-0.108
	Previous surgery	-2.936	0.905	-0.083	-3.244	0.001	-4.712	-1.159

	Duration of symptoms	-0.338	0.160	-0.060	-2.107	0.035	-0.652	-0.023
	Comorbidities	-1.601	0.322	-0.134	-4.979	0.000	-2.232	-0.970
	Baseline MSK-HQ score	0.562	0.031	0.489	18.157	0.000	0.502	0.623
	Previous pain episodes	-1.559	0.217	-0.208	-7.174	0.000	-1.986	-1.133

Variables of sex, age and pain site were removed from the model due to not meeting statistical parameters (removed if p>0.1). In model version d with these variables removed ANOVA showed a large statistically significant F ratio (101.598, p<0.000) showing a good fit to the data. The modified FOTO model had strong predictive power in this UK MSK community and primary care dataset. Adjusted R^2 was 0.438 meaning that 44% of the variation in MSK-HQ outcome at 6 months could be explained by the model/baseline factors (**Table 6-5**). Assumptions of normality and homoscedasticity were met and non-collinearity was satisfied. **Figure 6-1** and **6-2** show normally distributed error terms. **Figure 6-3** shows that there is no relationship between the error terms, meeting assumptions of homoscedasticity. Model power remained with EQ5D5L used as an alternative PROM (adjusted R2 0.435) (see **Table 6-6**).

Figure 6-1 Modified FOTO Model; Histogram showing normal distribution of error terms

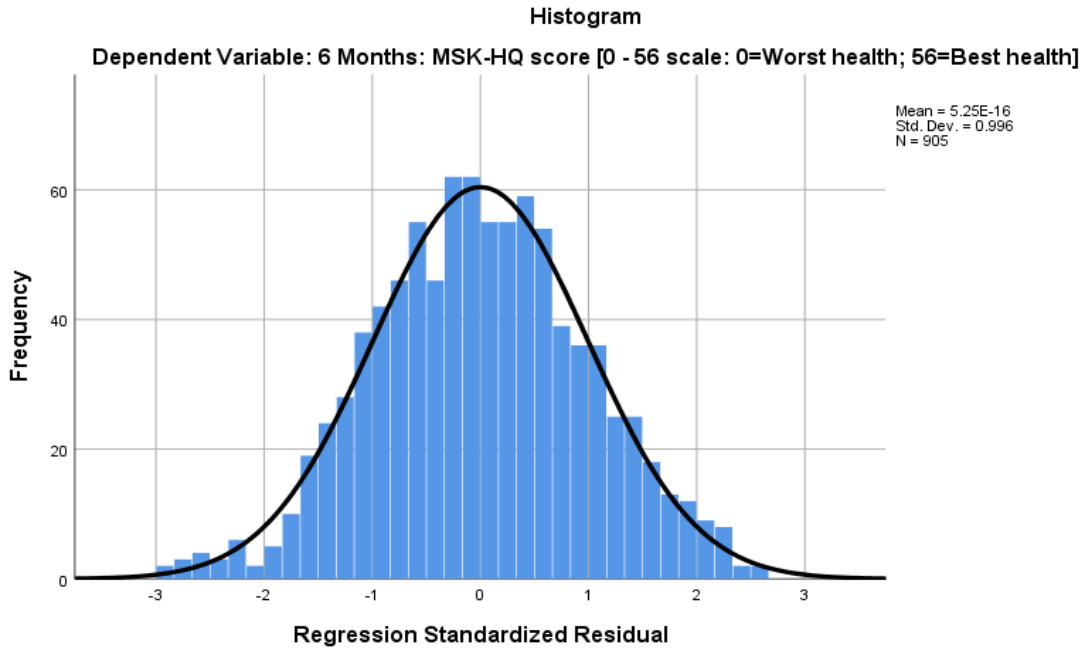


Figure 6-2 Modified FOTO Model; Plot of regression standardised residuals

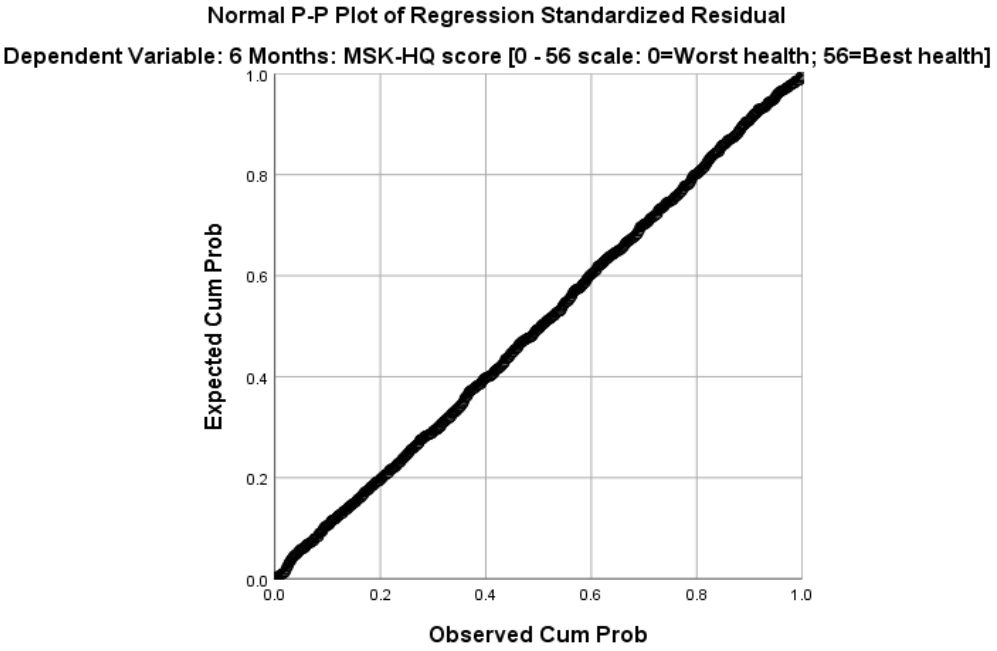


Figure 6-3 Modified FOTO Model; Scatterplot showing independence of error terms

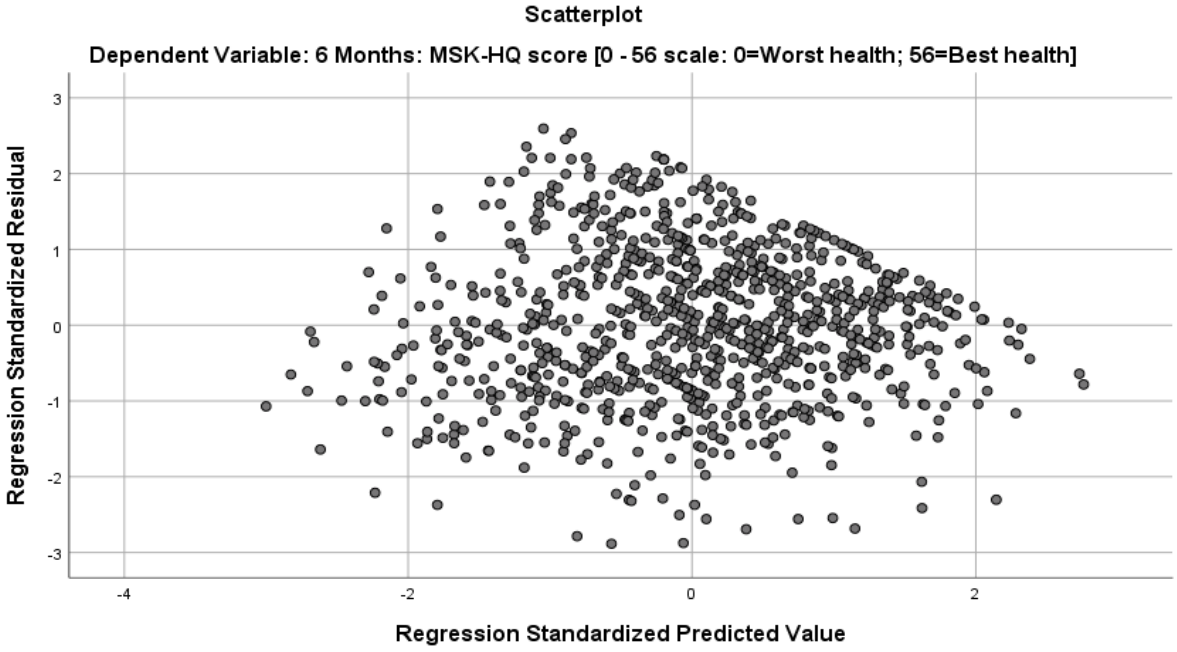


Table 6-6 EQ5D5L Modified FOTO Backward Stepwise Model; Summary and coefficients (model f)

Model Summary[g] (n=948)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
a.	.665[a]	0.442	0.437	0.1725838	0.442	74.362	10	937	0.000	
b.	.665[b]	0.442	0.437	0.1725405	0.000	0.529	1	937	0.467	
c.	.664[c]	0.441	0.437	0.1725614	-0.001	1.227	1	938	0.268	
d.	.664[d]	0.440	0.436	0.1726220	-0.001	1.661	1	939	0.198	
e.	.663[e]	0.439	0.436	0.1727236	-0.001	2.108	1	940	0.147	
f.	.662[f]	0.438	0.435	0.1728494	-0.001	2.373	1	941	0.124	1.959

a. Predictors: (Constant), Age, Sex, Physical activities, Baseline: Duration of symptoms, Pain site, Previous surgery, Baseline EQ5D score, Comorbidities, Previous pain episodes, Current paid employment

b. Predictors: (Constant), Age, Physical activities, Baseline: Duration of symptoms, Pain site, Previous surgery, Baseline EQ5D score, Comorbidities, Previous pain episodes, Current paid employment

c. Predictors: (Constant), Age, Physical activities, Duration of symptoms, Previous surgery, Baseline EQ5D score, Comorbidities, Previous pain episodes, Current paid employment

d. Predictors: (Constant), Age, Physical activities, Duration of symptoms, Baseline EQ5D score, Comorbidities, Previous pain episodes, Current paid employment

e. Predictors: (Constant), Age, Duration of symptoms, Baseline EQ5D score, Comorbidities, Previous pain episodes, Current paid employment

f. Predictors: (Constant), Duration of symptoms, Baseline EQ5D score, Comorbidities, Previous pain episodes, Current paid employment

g. Dependent Variable: 6 Month: EQ5D score

Coefficients

Model	Unstandardized Coefficients	Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B

		B	Std. Error	Beta			Lower Bound	Upper Bound
f	(Constant)	0.510	0.025		20.699	0.000	0.462	0.559
	Duration of symptoms	-0.009	0.003	-0.083	-2.969	0.003	-0.015	-0.003
	Current paid employment	0.039	0.012	0.084	3.289	0.001	0.016	0.062
	Baseline EQ5D score	0.508	0.026	0.508	19.592	0.000	0.457	0.559
	Comorbidities	-0.033	0.006	-0.139	-5.212	0.000	-0.045	-0.021
	Previous pain episodes	-0.024	0.004	-0.163	-5.818	0.000	-0.032	-0.016

Sensitivity analysis showed that all variables that were predictive in OLS models remained significantly predictive in mixed RE models. Fixed effects/coefficients are shown in **Table 6-7**. The Intra-class correlation coefficient (ICC) was 0.017, indicating that the proportion of variance in MSK-HQ score explained by GP practices (n=24 practices) was 0.017 or 1.7%.

Table 6-7: STATA Modified FOTO Mixed Effects Model Coefficients (Fixed Effects)

Variables	B Coefficient	Standard. Error	z	P>z	95% Confidence Interval for B	
Dependent Variable: 6 month MSK-HQ score					Lower bound	Upper bound
Age	0.007	0.028	0.26	0.792	-0.047	0.061
Sex (female)	0.098	0.595	0.16	0.869	-1.069	1.264
Duration of symptoms	-0.336	0.159	-2.11	**0.035**	-0.648	-0.024
Pain site (single site)	1.613	1.044	1.54	0.122	-0.434	3.659
Physical activity	-0.359	0.126	-2.84	**0.004**	-0.606	-0.111
Previous surgery	-2.656	0.900	-2.95	**0.003**	-4.419	-0.892
Baseline MSK-HQ score	0.562	0.031	17.93	**0.000**	0.500	0.623
Previous pain episode	-1.480	0.217	-6.81	**0.000**	-1.906	-1.054
Comorbidities	-1.477	0.323	-4.58	**0.000**	-2.110	-0.845
Current paid employment	2.198	0.759	2.90	**0.004**	0.711	3.685
Constant	27.77	2.563	10.83	**0.000**	22.75	32.79

Mixed-effects ML regression, Number of observations = 905
Group variable: Practice Code, Number of groups = 24

Modified NPROMS Model

896 of the 1211 STarT MSK Trial patients (74%) had complete data for available NPROMS variables. The model summary is shown in **Table 6-8** alongside the standardised coefficients for the final backward stepwise model (modified NPROMS model version e) with all redundant/non-significant variables removed.

Table 6-8 Modified NPROMS Backwards Stepwise Model: Model summary & coefficients

Model Summary[e] (n=896)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
a.	.649[a]	0.421	0.415	8.869	0.421	64.367	10	885	0.000	
b.	.649[b]	0.421	0.415	8.867	0.000	0.601	1	885	0.438	
c.	.648[c]	0.420	0.415	8.867	-0.001	0.994	1	886	0.319	
d.	.647[d]	0.419	0.414	8.872	-0.001	2.121	1	887	0.146	1.918

a. Predictors: (Constant), Baseline EQ5D score, Age, Sex, Duration of symptoms, Pain site, IMD decile, Previous surgery, Live alone, Comorbidities, Baseline MSK-HQ score

b. Predictors: (Constant), Baseline EQ5D score, Age, Duration of symptoms, Pain site, IMD decile, Previous surgery, Live alone, Comorbidities, Baseline MSK-HQ score

c. Predictors: (Constant), Baseline EQ5D score, Age, Duration of symptoms, Pain Site, Previous surgery, Live alone, Comorbidities, Baseline MSK-HQ score

d. Predictors: (Constant), Baseline EQ5D score, Duration of symptoms, Pain Site, Previous surgery, Live alone, Comorbidities, Baseline MSK-HQ score

e. Dependent Variable: 6 Months: MSK-HQ score [0 - 56 scale: 0=Worst health; 56=Best health]

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
d	(Constant)	26.411	1.274		20.736	0.000	23.911	28.911
	Pain site	-2.957	1.041	-0.074	-2.840	0.005	-5.001	-0.913
	Previous surgery	-3.658	0.922	-0.104	-3.967	0.000	-5.468	-1.849
	Duration of symptoms	-0.755	0.146	-0.136	-5.157	0.000	-1.043	-0.468
	Comorbidities	-1.538	0.325	-0.129	-4.730	0.000	-2.175	-0.900

	Baseline MSK-HQ score	0.392	0.045	0.342	8.639	0.000	0.303	0.481
	Live alone	-1.440	0.795	-0.047	-1.811	0.070	-3.001	0.120
	Baseline: EQ5D score	10.007	2.023	0.198	4.946	0.000	6.036	13.979

Variables of IMD, sex, and age were removed from the model due to not meeting statistical parameters (removed if p>0.1). ANOVA with these variables removed (model version d in **Table 6-8**) showed a large statistically significant F ratio (91.349, p<0.000) indicating a good fit to the data. The modified NPROMs model had strong predictive power in this UK MSK community and primary care dataset. Adjusted R^2 was 0.414 meaning that 41% of the variation in MSK-HQ outcome at 6 months could be explained by the model/baseline factors. Assumptions of normality and homoscedasticity were met and non-collinearity was satisfied (see **Figures 6-4, 6-5, 6-6**). **Figure 6-4** and **6-5** show normally distributed error terms. **Figure 6-6** shows that there is no relationship between the error terms, meeting assumptions of homoscedasticity. Model power remained with EQ5D5L used as an alternative PROM (adjusted R2 0.419) (see **Table 6-9**).

Figure 6-4 Modified NPROMS Model; Histogram showing normal distribution of error terms

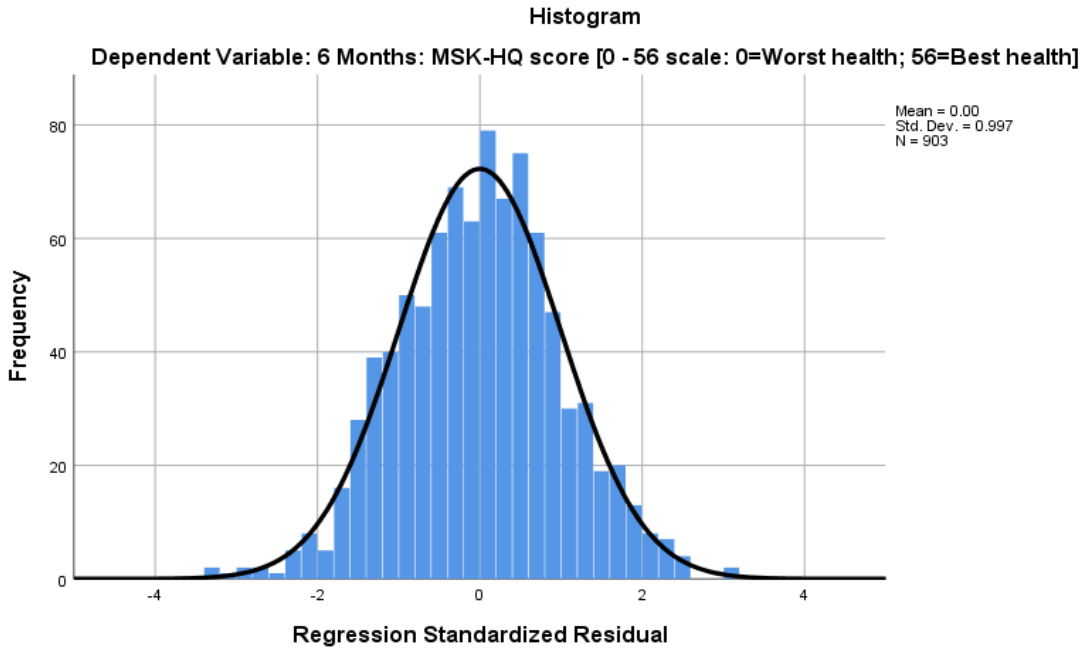


Figure 6-5 Modified NPROMS Model; Plot of regression standardised residuals

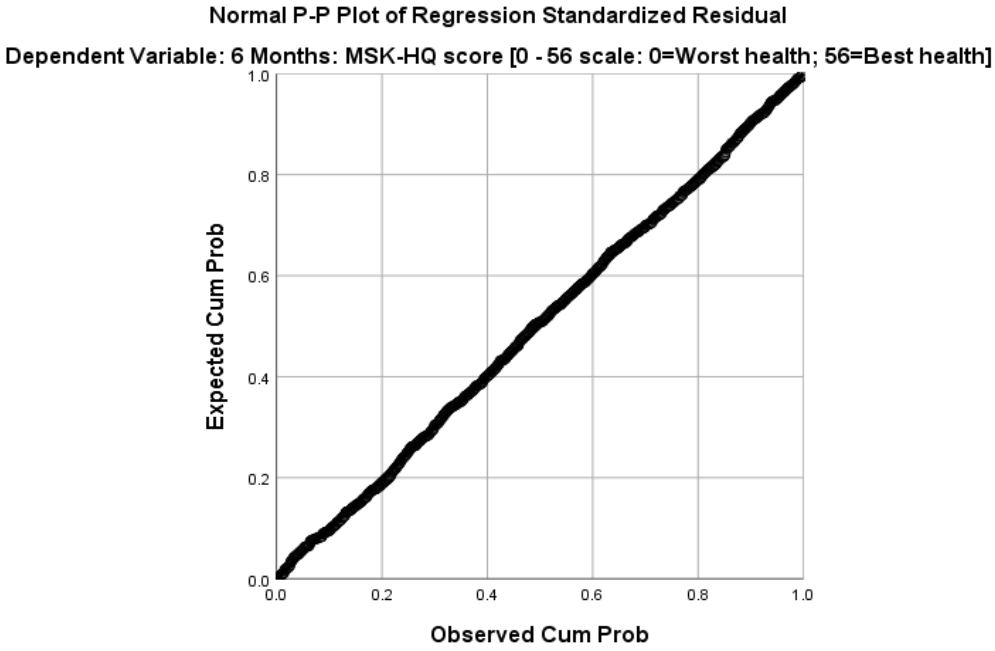


Figure 6-6 Modified NPROMS Model; Scatterplot showing independence of error terms

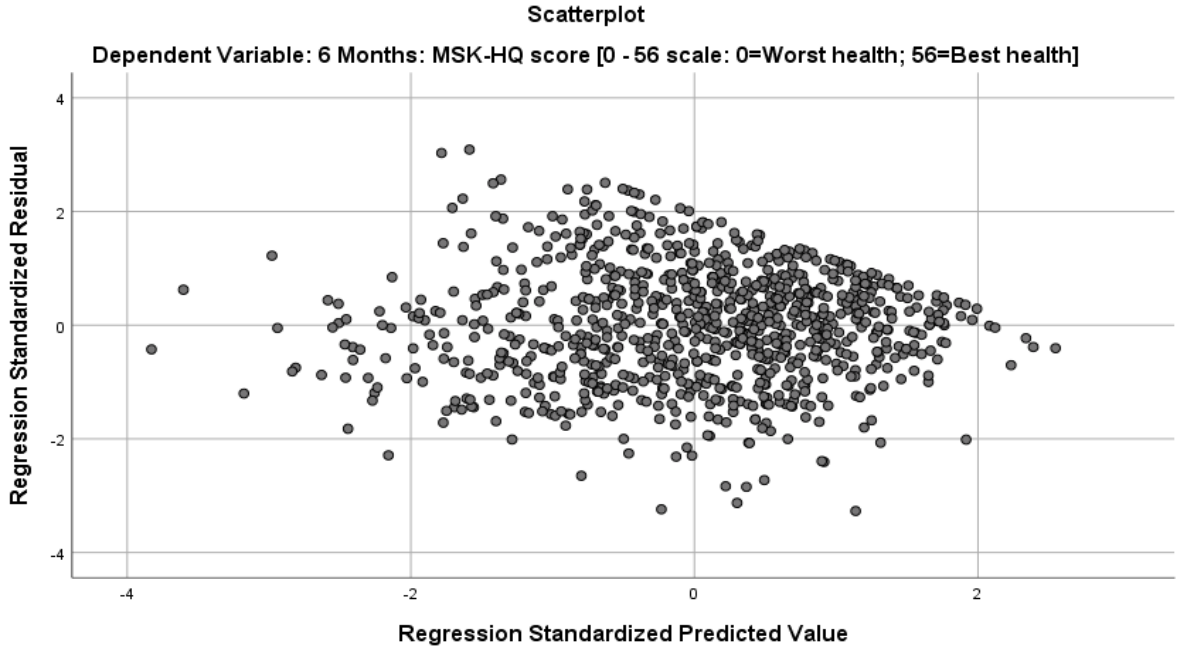


Table 6-9 EQ5D Modified NPROMS Backward Stepwise Model Summary and Coefficients (model d)

Model Summary[e] (N=968)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
a.	.652[a]	0.425	0.420	0.1771196	0.425	78.706	9	958	0.000	
b.	.652[b]	0.425	0.420	0.1770306	0.000	0.036	1	958	0.850	
c.	.652[c]	0.425	0.420	0.1770083	0.000	0.758	1	959	0.384	
d.	.650[d]	0.423	0.419	0.1771630	-0.002	2.681	1	960	0.102	1.946

a. Predictors: (Constant), Live alone, Pain site, IMD decile, Duration of symptoms, Sex, Previous surgery, Comorbidities, Age, Baseline EQ5D score

b. Predictors: (Constant), Live alone, Pain site, IMD decile, Duration of symptoms, Sex, Previous surgery, Comorbidities, Baseline EQ5D score

c. Predictors: (Constant), Live alone, Pain Site, IMD decile, Duration of symptoms, Previous surgery, Comorbidities, Baseline EQ5D score

d. Predictors: (Constant), Live alone, Pain site, Duration of symptoms, Previous surgery, Comorbidities, Baseline EQ5D score

e. Dependent Variable: 6 Month: EQ5D score

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
d	(Constant)	0.480	0.022		22.017	0.000	0.437	0.523
	Duration of symptoms	-0.016	0.003	-0.139	-5.508	0.000	-0.021	-0.010
	Pain site	-0.041	0.020	-0.051	-2.052	0.040	-0.080	-0.002
	Previous surgery	-0.032	0.017	-0.046	-1.827	0.068	-0.066	0.002
	Baseline EQ5D score	0.524	0.026	0.521	20.064	0.000	0.473	0.575

	Comorbidities	-0.036	0.006	-0.150	-5.753	0.000	-0.048	-0.024
	Live alone	-0.044	0.015	-0.072	-2.887	0.004	-0.074	-0.014

Sensitivity analysis showed that all variables that were predictive in OLS models remained significantly predictive in a mixed RE model. Fixed effects/coefficients are shown in **Table 6-10**. The Intra-class correlation coefficient (ICC) was 0.014, indicating that the proportion of variance in MSK-HQ score explained by GP practices (n=24 practices) is 0.014 or 1.4%.

Table 6-10: STATA Modified NPROMS Mixed Effects Model Coefficients

Variables	B Coefficients	Standard Error	z	P>z	95% Confidence Interval for B	
Dependent variable: 6 month MSK-HQ score					Lower bound	Upper bound
Age	-0.034	0.023	-1.47	0.142	-0.078	0.011
Sex	-0.501	0.613	-0.82	0.414	-1.702	0.701
Duration of symptoms	-0.765	0.146	-5.25	**0.000**	-1.050	-0.479
Previous surgery	-3.387	0.917	-3.69	**0.000**	-5.185	-1.589
Baseline MSK-HQ score	0.401	0.046	8.78	**0.000**	0.311	0.490
Comorbidities	-1.352	0.332	-4.07	**0.000**	-2.002	-0.701
IMD Decile	0.126	0.140	0.90	0.366	-0.148	0.401
Baseline EQ5D	9.730	2.024	4.81	**0.000**	5.762	13.70
Living alone	-1.122	0.808	-1.39	0.165	-2.706	0.461
Pain site	2.716	1.055	2.57	**0.010**	0.648	4.785
Constant	24.82	2.210	11.23	**0.000**	20.49	29.15

Mixed-effects ML regression, Number of observations = 896

Group variable: Practice Code, Number of groups = 24

Objective 2:

Keele Model 1: Evidence informed model

873 of the 1211 patients in the STarT MSK trial cohort (72%) had complete data for variables entered into the evidence informed Keele Model 1. The independent variable of 'distress' was not entered as a mental health variable (following recommendations in **Table 6-2**) due to high correlation of this variable with pain intensity (r=0.811) meaning that it looks to be measuring the same construct and does not meet assumptions of independence. Variables with very strong evidence made up 'model version a', variables with strong evidence 'model version b' and variables with moderate evidence 'model version c' (see **Table 6-11** showing model summary).

Assumptions of normality and homoscedasticity were met and non-collinearity was satisfied (**See Figures 6-7, 6-8, 6-9**). **Figure 6-7** and **6-8** show normally distributed error terms. **Figure 6-9** shows that there is no relationship between the error terms, meeting assumptions of homoscedasticity.

Results show that adding in the variables with 'moderate evidence' gave only a very small improvement to model power (R^2 change of 0.006) and therefore model version b using the variables with strong evidence to support their inclusion would be the preferred model in practice. This model explained 41% of the variation in 6-month MSK-HQ outcome and was statistically more predictive than model version a.

Table 6-11 Keele Model 1: Evidence informed model summary and coefficients (model c)

Model Summary[d] (n=873)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
a.	.598[a]	0.357	0.355	9.276	0.357	161.074	3	869	0.000	
b.	.646[b]	0.417	0.411	8.863	0.059	17.585	5	864	0.000	
c.	.650[c]	0.422	0.415	8.836	0.006	2.788	3	861	0.040	1.959

a. Predictors: (Constant), Baseline Comorbidities, Age, Baseline: MSK-HQ score

b. Predictors: (Constant), Baseline Comorbidities, Age (years), Baseline: MSK-HQ score, Previous surgery, IMD decile, Duration of symptoms, Current paid employment, Baseline EQ5D score

c. Predictors: (Constant), Comorbidities, Age, Baseline: MSK-HQ score, Previous surgery, IMD decile, Duration of symptoms, Current paid employment, Baseline EQ5D score, Sex, Pain site, Pain intensity

d. Dependent Variable: 6 Months: MSK-HQ score [0 - 56 scale: 0=Worst health; 56=Best health]

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
c	(Constant)	21.860	3.067		7.127	0.000	15.840	27.880
	Baseline: MSK-HQ score	0.435	0.053	0.380	8.156	0.000	0.330	0.539
	Age	-0.003	0.029	-0.003	-0.091	0.927	-0.060	0.054
	Comorbidities	-1.315	0.342	-0.110	-3.841	0.000	-1.987	-0.643
	Previous surgery	-3.293	0.938	-0.094	-3.512	0.000	-5.134	-1.453
	IMD decile	0.181	0.137	0.035	1.321	0.187	-0.088	0.449
	Duration of symptoms	-0.819	0.148	-0.148	-5.526	0.000	-1.109	-0.528
	Current paid employment	2.134	0.797	0.092	2.679	0.008	0.570	3.698
	Baseline: EQ5D score	9.408	2.061	0.186	4.564	0.000	5.363	13.454

Sex	-0.396	0.623	-0.017	-0.636	0.525	-1.619	0.827
Pain site	-2.478	1.064	-0.062	-2.328	0.020	-4.567	-0.389
Pain intensity	0.247	0.182	0.048	1.359	0.174	-0.110	0.603

a; variables with ‘very strong’ evidence

b; variables with ‘very strong’ and ‘strong’ evidence

c; variables with ‘very strong’ evidence, ‘strong’ evidence and ‘moderate evidence’.

Figure 6-7 Keele Model 1; Histogram showing normal distribution of error terms

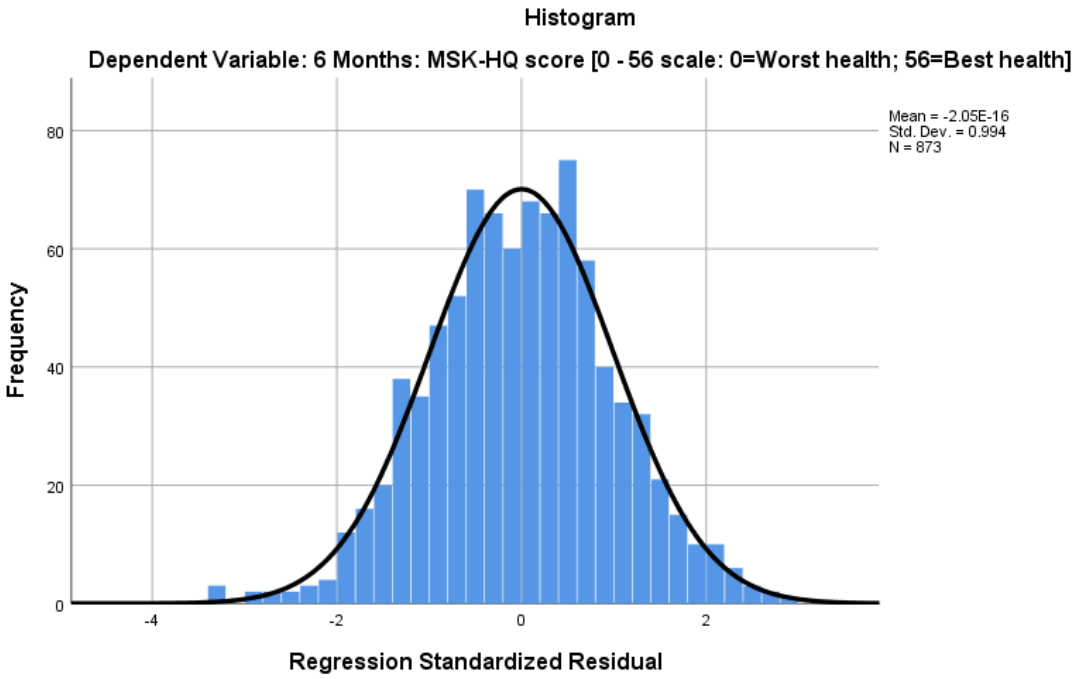


Figure 6-8 Keele Model 1; Normal P-Plot of regression standardised residual

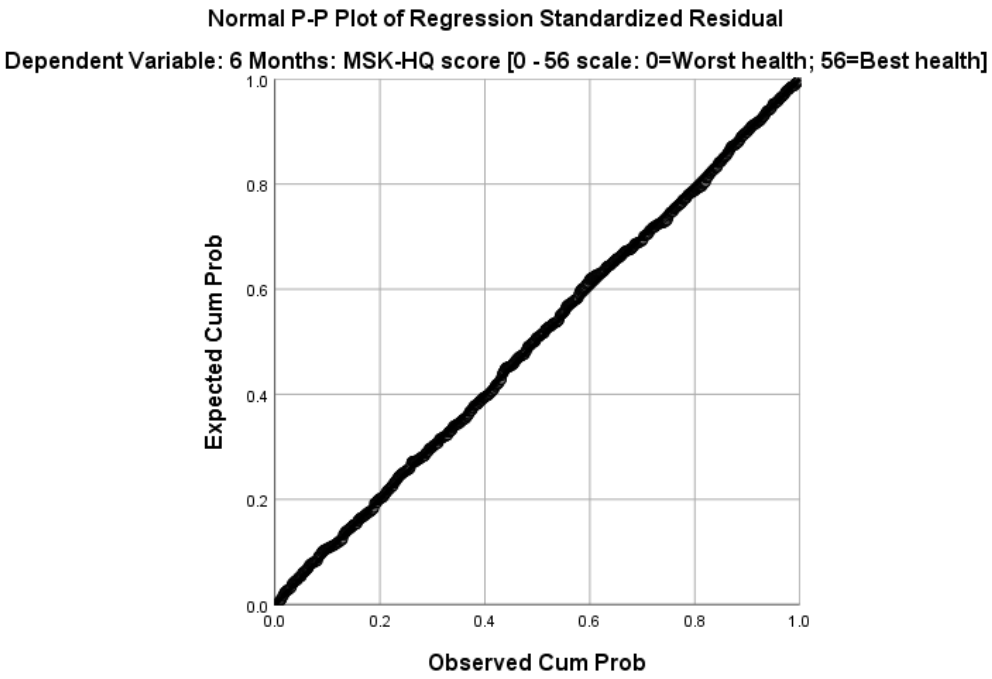
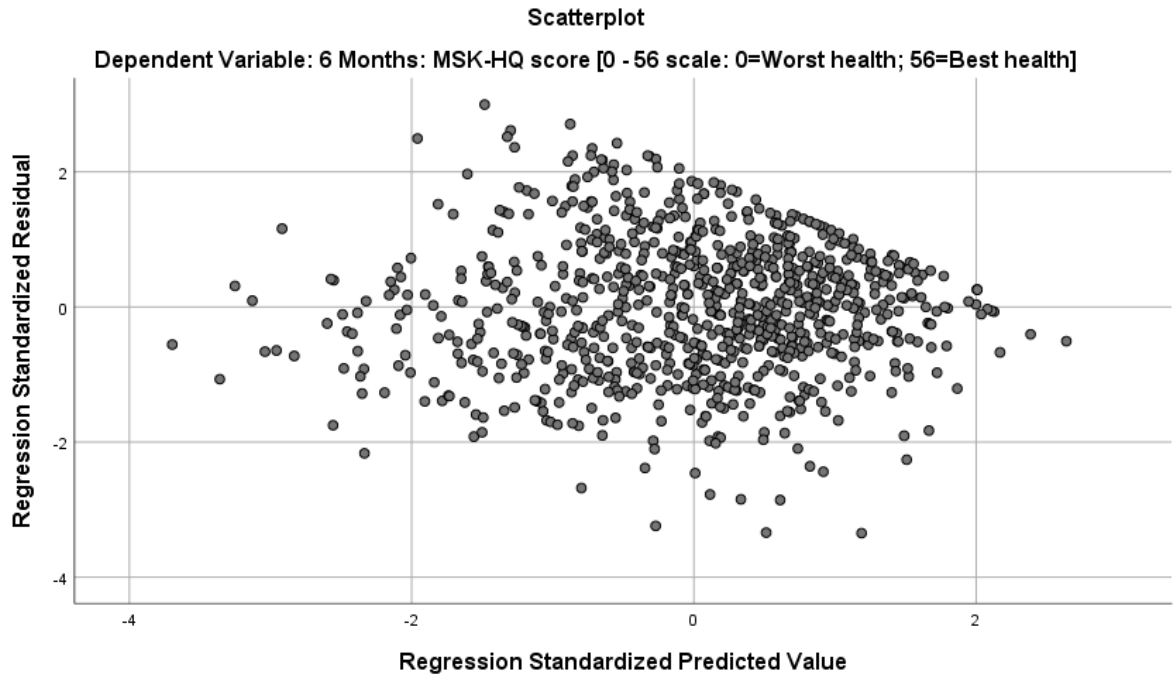


Figure 6-9 Keele Model 1: Scatterplot showing independence of error terms



Objective 3:

Keele Model 2: Statistically informed model

17 variables were available for the model giving 850 patients with complete data (70.19%). The baseline variable of distress (continuous) was removed due to being too highly correlated with baseline pain intensity (r=0.819) and the variables of 'performance at work' and 'time off work' were removed due to reduced numbers with these variables complete (including these variables excluded all participants not in paid employment and led to only 34% with complete data).

The model summary is shown in **Table 6-12**. Assumptions of normality and homoscedasticity were met and non-collinearity was satisfied. **Figure 6-10** and **6-11** show normally distributed error terms. **Figure 6-12** shows that there is no relationship between the error terms, meeting assumptions of homoscedasticity.

Table 6-12 Keele Model 2 Summary: Statistically informed model summary and coefficients (forward stepwise model demonstrating additional model strength as variables added)

Model Summary[j] (n=850)										
					Change Statistics					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
a.	.564[a]	0.318	0.318	9.500	0.318	396.228	1	848	0.000	
b.	.622[b]	0.387	0.386	9.012	0.069	95.343	1	847	0.000	
c.	.643[c]	0.414	0.412	8.820	0.027	38.289	1	846	0.000	
d.	.656[d]	0.430	0.427	8.703	0.016	23.765	1	845	0.000	
e.	.664[e]	0.442	0.438	8.620	0.012	17.445	1	844	0.000	
f.	.670[f]	0.449	0.445	8.568	0.007	11.353	1	843	0.001	
g.	.675[g]	0.455	0.450	8.525	0.006	9.361	1	842	0.002	
h.	.679[h]	0.461	0.455	8.487	0.006	8.660	1	841	0.003	
i.	.681[i]	0.463	0.458	8.470	0.003	4.425	1	840	0.036	1.927
a. Predictors: (Constant), Baseline: MSK-HQ score										
b. Predictors: (Constant), Baseline: MSK-HQ score, Baseline: Previous pain episodes										
c. Predictors: (Constant), Baseline: MSK-HQ score, Previous pain episodes, Health Literacy										
d. Predictors: (Constant), Baseline: MSK-HQ score, Previous pain episodes, Health Literacy, Comorbidities										
e. Predictors: (Constant), Baseline: MSK-HQ score, Previous pain episodes, Health Literacy, Comorbidities, Baseline EQ5D score										
f. Predictors: (Constant), Baseline: MSK-HQ score, Previous pain episodes, Health Literacy, Comorbidities, Baseline EQ5D score, Current paid employment										
g. Predictors: (Constant), Baseline: MSK-HQ score, Previous pain episodes, Health Literacy, Comorbidities, Baseline EQ5D score, Current paid employment, Previous surgery										
h. Predictors: (Constant), Baseline: MSK-HQ score, Previous pain episodes, Health Literacy, Comorbidities, Baseline EQ5D score, Current paid employment, Previous surgery, Physical activities										
i. Predictors: (Constant), Baseline: MSK-HQ score, Previous pain episodes, Health Literacy, Comorbidities, Baseline EQ5D score, Current paid employment, Previous surgery, Physical activities, Duration of symptoms										
j. Dependent Variable: 6 Months: MSK-HQ score [0 - 56 scale: 0=Worst health; 56=Best health]										
Coefficients										

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
i.	(Constant)	32.757	1.602		20.444	0.000	29.612	35.902
	Baseline: MSK-HQ score	0.389	0.045	0.340	8.671	0.000	0.301	0.477
	Previous pain episodes	-1.524	0.219	-0.206	-6.965	0.000	-1.954	-1.095
	Health Literacy	-2.045	0.433	-0.127	-4.720	0.000	-2.896	-1.195
	Comorbidities	-1.098	0.333	-0.092	-3.295	0.001	-1.752	-0.444
	Baseline: EQ5D score	8.131	1.982	0.161	4.103	0.000	4.241	12.022
	Current paid employment	2.012	0.610	0.087	3.295	0.001	0.813	3.210
	Previous surgery	-2.625	0.920	-0.074	-2.854	0.004	-4.430	-0.820
	Physical activities	-0.376	0.128	-0.076	-2.933	0.003	-0.627	-0.124
	Duration of symptoms	-0.339	0.161	-0.061	-2.103	0.036	-0.655	-0.023

Figure 6-10 Keele Model 2; Histogram showing normal distribution of error terms

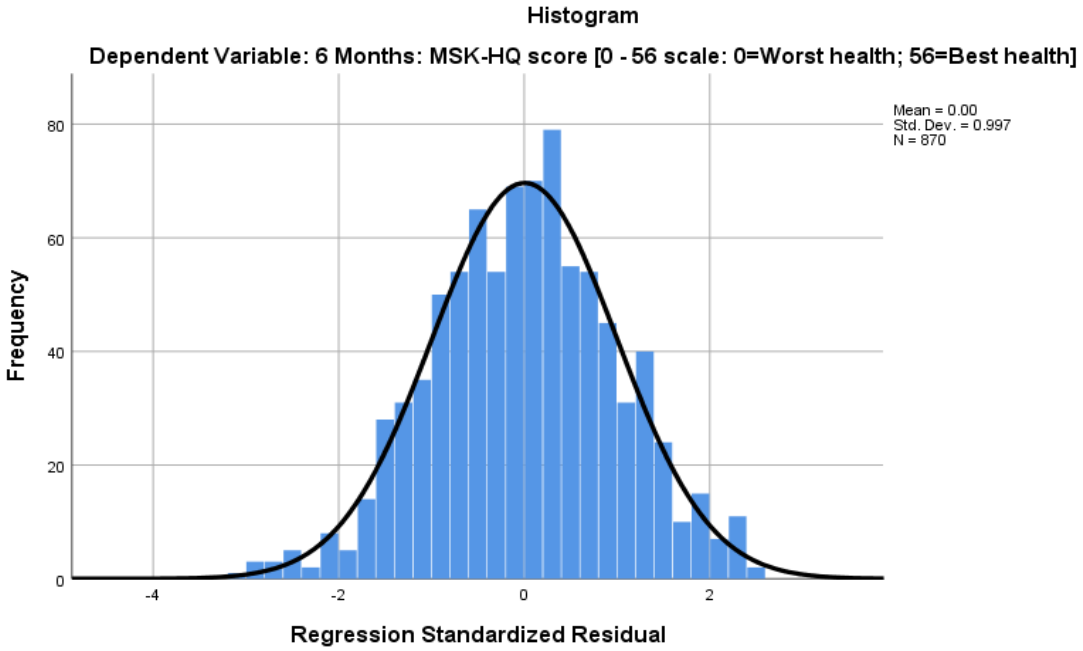


Figure 6-11 Keele Model 2; Normal P-Plot of Regression Standardised Residual

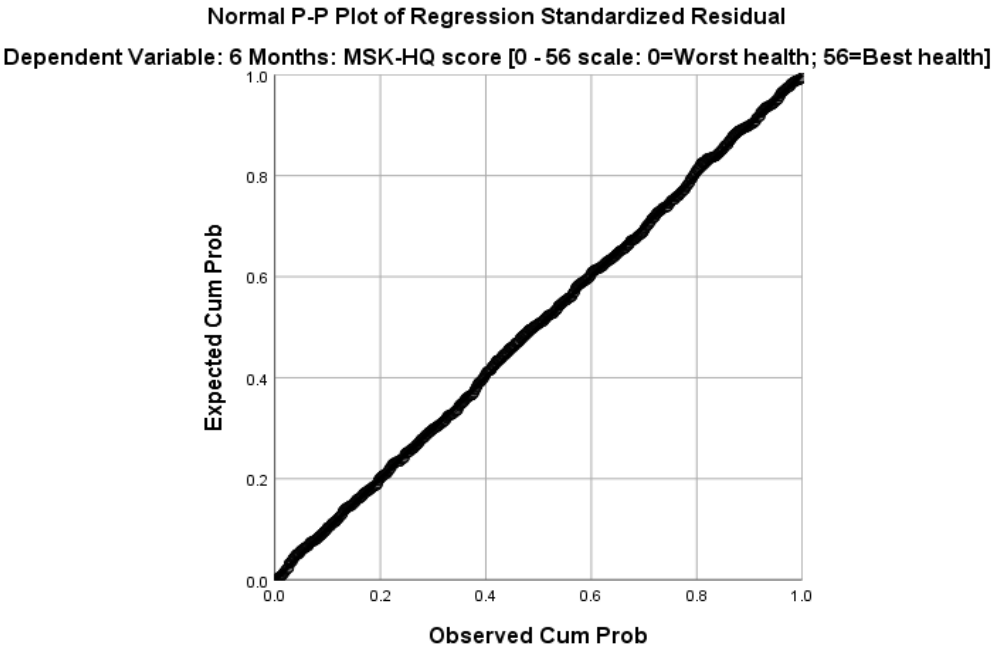
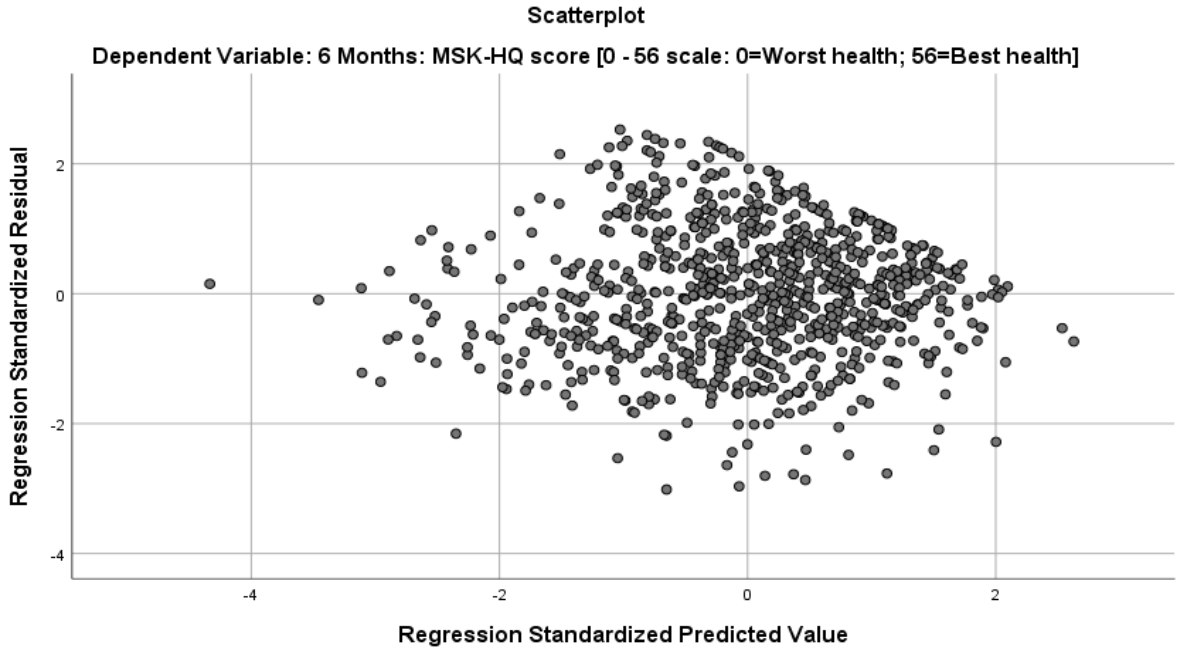


Figure 6-12 Keele Model 2; Scatterplot showing independence of error terms



The statistically informed Keele Model 2 considered 17 variables for entry to the model and resulted in 9 variables being retained in the final model. A forward stepwise approach was used to clearly demonstrate the additional model strength with addition of each variable that met statistical parameters. The variables of age, gender, IMD, pain intensity, pain site, living alone, confidence managing pain and TSK score were removed due to not meeting statistical parameters. Of the 9 variables retained, the most predictive were: baseline MSK-HQ score, previous pain episodes, health literacy, comorbidities, baseline EQ5D5L score, and current paid employment respectively (see **Table 6-12**). These factors alone explained 45% of the variation in outcome.

Discussion

41-46% of the change in the primary outcome score (MSK-HQ) from baseline to 6 months could be predicted by pre-defined baseline factors for MSK patients treated in this primary/community care setting. Unmeasured/unknown factors related to the patient, clinical performance and error make up the remainder of the PROM score change for each patient (Lutz et al, 2020).

Objective 1 Summary

Both the modified FOTO and the modified NPROMS case-mix models were highly predictive in this UK community and primary care StarT MSK Trial dataset. Limitations of this analysis include the size of the dataset (n=1211) comparative to the datasets used in both the FOTO (n= 341,642 lumbar) and NPROMs model (39,404-47,392 hips, 45,773-54,062 knees) development papers (Deutscher et al, 2018, DoH, 2012, NHSE 2013). The categories included to collect/report variable information and the dependent variable of PROM outcome were also not standardised between models meaning that neither the FOTO nor NPROMS model was exactly replicated.

Of the two models tested, the model based on variables within the FOTO case-mix methodology (Deutscher et al, 2018) was slightly more predictive (R^2 0.44) compared to the modified NPROMS model (R^2 0.41). Both models predicted the MSK-HQ outcome within the STarT MSK Trial dataset better than they predicted functional outcomes in the developmental papers (FOTO R^2 0.37 (Deutscher et al, 2018) and NPROMS R^2 0.23-0.30 (DoH, 2012)). The predictive ability also remained high when an alternative PROM (EQ5D5L) was used (R^2 0.44, R^2 0.42 respectively). Both models removed variables of age and sex, with the modified FOTO model additionally removing pain site, and the modified NPROMS model additionally removing IMD. Medication at baseline could not be added to the FOTO model in this study due to not being available at this time within the STarT MSK Trial dataset, this may further improve model fit.

The FOTO case-mix adjustment model was developed in US outpatient rehabilitation clinics similar to a UK community/primary care setting and the model was reported by Deutscher et al (2018) for those with low back pain which was also the largest pain site group within the STarT MSK trial dataset (37.7%) which may be why this model performed slightly better than the UK model developed for knee/hip surgical patients. Sensitivity analyses using mixed RE models showed ICC estimates of; 0.017, and 0.014, for the modified FOTO and modified NPROMs models respectively, indicating the proportion of variance in MSK-HQ scores explained by GP practices

(n=24 practices). Understanding how practices differ with regards to PROM outcomes is an important first step in further investigating the differences in outcome performance between practices (Lorah, 2018), and whether this nesting of patients within practices should be adjusted for.

Objective 2 Summary

Keele Model 1 which was developed using all variables with strong supporting evidence (model version b in **Table 6-7**). This model included 8 variables and explained 41% of the total treatment outcome at 6 months. This evidence informed model was therefore as strong as the modified NPROMS predictive model, but slightly less predictive than the modified FOTO model. This remained the case with addition of variables with 'moderate evidence'.

Objective 3 Summary

The final model was developed using all available variables within the STarT MSK Trial dataset rather than following the recommended guidelines around using available literature to inform variable selection (Krumholz et al, 2006). This model (Keele Model 2) retained 9 independent variables. The most predictive of these were baseline MSK-HQ score, previous pain episodes, health literacy, comorbidities, baseline EQ5D5L, and current paid employment respectively. The 9-variable model explained 46% of the total variation in treatment outcome and therefore as expected was the strongest of all models. If this was reduced to 6 variables the model still explained 45% of total variation and with 4 variables 43%. Due to the sample size however, we did not split the sample to internally validate these results. These findings suggest that as a minimum 'parsimonious model', the variables of baseline MSK-HQ, previous pain episodes, health literacy and comorbidities should be included within a case-mix model for use in UK primary/community care. This is interesting as health literacy was not considered within the development of existing case-mix models and there is limited literature in this area, and previous

pain episodes replaced previous treatment (used by the FOTO team) but may be a better fit to the data.

Summary

In summary all models demonstrated strong predictive ability ranging from 41-46%. This study provides external validation to the FOTO (Deutscher et al, 2018) and the NPROMS (DoH, 2012, NHSE 2013) case-mix adjustment models taking modifications into account. These models remained highly predictive with use of an alternative PROM (EQ5D5L) providing further support for the strength of the modified US FOTO and UK NPROMS models for use in case-mix adjustment of MSK PROM data.

The modified FOTO model was the strongest of existing models. Variables retained within the model were: current paid employment, days in last week doing moderate physical activities, previous surgery, symptom duration, comorbidities, baseline MSK-HQ score, and previous pain episodes. These seven variables are feasible for collection in routine clinical practice in the UK health system. We therefore recommend that this model based on many years of development and refinement, is used as the preferential model for adjusting UK primary/community healthcare data at this time. Additional variables within the FOTO model of pain site, sex and age should also be included and further evaluated in a larger dataset alongside use of medication at baseline. Future model development/testing would also be beneficial to validate findings from our statistically informed Keele Model 2 in a larger dataset to see if the variables of baseline MSK-HQ score, health literacy, previous pain episodes and comorbidities remained highly predictive in a UK community/primary care setting, and to test newly emerging predictors such as 'recovery expectations' (van der Gaag et al, 2020) which were not available in this dataset.

Limitations

A significant limitation of this study is that the data collection approach was not standardised across the STarT MSK cluster RCT study and the developmental studies detailing the FOTO and NPROMS models. Direct comparison between these studies is therefore limited due to the

differences in primary outcome, baseline variable categories, and timing and method of collection. This study does however provide some evidence to show that the models developed with alternative MSK PROM tools do seem to be transferable across other MSK functional status measures as both models performed well in predicting the primary MSK-HQ outcome and the EQ5D5L outcome within this patient cohort.

Another limitation of this study was not splitting the data into a training dataset and a predictive validation sample in order to internally validate the Keele developmental models. It is therefore possible that the statistically informed (Keele Model 2) model over-fits the data and would not be as effective outside of this STarT MSK Trial dataset thus reducing generalisability. It is therefore important to interpret the results of this model with consideration to these limitations as in another sample the adjusted R^2 value may not be as high.

The reduced diversity of this study population made the variable of 'ethnicity' untestable due to the minimal variation across categories. This would need to be analysed fully in a more diverse and representative population to see whether it's inclusion within case-mix modelling is important. Nuttall et al (2015) reported that patients recorded as Asian or Black had on average worse outcomes than those whose ethnicity was recorded as White within the NPROMs Oxford Knee Score data.

Completion rates for variable data were high within this dataset. For testing the modified FOTO model 75% of patients had complete data and so were included within the analysis, 74% for the modified NPROMS, 72% for Keele model 1 and 70% for Keele model 2. Patient selection/non-response bias becomes a critical issue if response rates fall below 70% (Prince, 2012), it is therefore not a significant issue within this study.

Sensitivity analyses using mixed/multi-level models accounting for clustering of patients between GP practices were conducted when exploring the use of a modified FOTO and modified NPROMS

model. These analyses showed 1.7% and 1.4% additional predictive power respectively when taking random effects (GP practices) into account. Compared with other similar studies the percentage of additional predictive power explained by these RE models was very low, with Yen et al (2015) reporting 11-12% of additional between patient differences being explained by use of a RE model compared to a FE model. This sensitivity analysis shows that use of a mixed/RE model had little impact on model predictive power compared to the OLS model. This may be due to all fixed effects being at the patient level rather than at a higher practice-based level.

Conclusion

In this chapter four case-mix adjustment models have been presented and compared with regards to their ability to explain variation in MSK-HQ outcomes at 6 months in a primary/community care patient population. Of these models the modified US FOTO model is recommended for use in UK community/primary care. Further research is needed to capture prospective routine data in this UK primary/community MSK setting on a large scale to further analyse the performance of this case-mix adjustment model and to assess its ability to benchmark performance and identify positive and negative outliers across MSK providers. This has never been of greater importance in the UK and internationally as we move out of the COVID-19 pandemic and try to understand the impact of the pandemic on MSK patients, national health services and outcome variation, and also identify where novel system changes are helpful in restoring capabilities for the future.

7 Chapter 7: Recommendations for future MSK benchmarking in community/primary care.

Chapter 7 looked to formulate recommendations for the development of MSK benchmarking in community/primary care MSK services, and to provide a framework for a national MSK evaluation/audit, with a detailed proposal for future research.

Abstract

Introduction: High quality data on service performance is essential in healthcare to evidence efficacy, efficiency, and value. There remains a paucity of publicly reported data in community and primary care MSK services. There is also a lack of guidance on which metrics MSK services should be collecting and reporting, and how this data could be used to directly improve patient outcomes and experiences and improve value through an MSK episode of care.

Method: A narrative review of the evidence around benchmarking MSK services was undertaken with a focus on how to develop routine data collection within community and primary care settings, and how to develop benchmarking capabilities for the future, looking towards a national MSK audit. This chapter brings together the findings from all of the previous chapters to create a summary recommendation. It is intended that this work forms the basis of a post-doctoral project/funding application to develop a national MSK data collection in community/primary care.

Recommendations: Taking the two systematic reviews (**Chapter 2 and 3**), one umbrella review (**Chapter 4**), national guidelines/policy and consensus study findings (**Chapter 5**), and case-mix adjustment analyses (**Chapter 6**) alongside emerging policy within the Best MSK Health programme into account, we have developed a recommendation on what to include within a minimum 'core outcome set'. The core set includes: demographics; **age, sex, ethnicity,** clinical factors; **pain site, comorbidities, duration of symptoms, previous surgery, previous pain**

episodes, PROM; **MSK-HQ**, PREM; **National MSK PREM** which will include domains of; **access/waiting time, understanding and engagement, overall experience (FFT), ideas to make us better, shared decision making, confidence in clinician, given sufficient information, convenience/timeliness, needs met, care planning**. Alongside recommendations on what to collect, we recommend the use of methods used by the National PROMs NHS Digital team for case-mix adjustment and outlier identification, alongside consideration to how this data could be integrated and reported as part of the Community Services Dataset (CSDS) and new Community Model Health System developed by NHS Improvement to ensure data reported contains system, process, and quality indicators for use in benchmarking MSK services.

Conclusions: Capturing high quality MSK data in a standardised, consistent, and sustainable way is a significant challenge. Policy holders, commissioners, managers and clinicians need to be realistic with expectations, and take time to explore barriers to implementation such as; funding, digital infrastructure/intra-operability, data sharing/governance, digital literacy, and local and national leadership. The development of infrastructure, governance, advanced analytics, and clear leadership, along with further exploration of barriers and enablers is warranted. If funded as a pilot study working alongside NHS England and the Best MSK Health Collaborative, this could identify optimal solutions to pave the way for a future sustainable national MSK audit programme and provide a model of continuous improvement in care quality for patients living with MSK conditions.

Introduction

How do patients with MSK conditions know where they can receive high quality care in their local region? How do MSK clinicians demonstrate the care outcomes they provide to their patients? How do managers and providers show that their MSK service delivers high value care? How do commissioners/funders of care know which providers are performing best? We know the answer to all of these questions is data (Hibbert et al 2020), but what data metrics should we use and

how do we benchmark MSK services in a way that facilitates and encourages healthy quality improvement discussions and not just competitive rankings? In this Chapter a set of recommendations are made that seek to answer these important questions, bringing together findings from previous Chapters.

Measuring health service performance involves collecting data in a number of key areas (Hibbert et al, 2020), including: effectiveness and appropriateness of care (for example effectiveness of an intra-articular joint injection measured using Patient Reported Outcome Measures (PROMs); or appropriateness of a lumbar spine MRI measured against set standards such as NICE Guidelines (NICE, 2016)); safety of care (for example providing information/advice such as safety netting patients for serious spinal pathology); efficiency of services (for example capturing number of healthcare visits for treatment of an MSK condition to a GP, Physiotherapist, or Secondary Care specialist (Burgess et al, 2020, **Chapter 4**)); and patient experience (for example use of tools such as the Friends and Family Test (NHS England, 2019, 2020) to ensure that patients are satisfied with the care received, or the CollaboRATE tool (Elwyn et al, 2013) to measure patient perceptions of involvement in decision-making about their care). All of these types of performance metrics are integral to ensuring an MSK service is delivering high quality care in the eyes of patients, clinicians, providers and commissioners/funders. Capturing this type of data routinely across an MSK episode of care for every patient however is challenging, particularly as benchmarking can only be achieved if it is captured in a standardised and transparent way.

Huge variation in the care and management of MSK conditions exist. An example from the Getting It Right First Time (GIRFT) programme showed a 25-fold variation in surgical site infections for MSK patients (Briggs, 2012). A recent primary care paper showed a 30-fold variation in GP MSK MRI requesting within a routine care cohort evaluation (Sajid et al, 2021), with significant downstream effects on costs and patient care. A third of GP MSK MRIs reported within this study were for lumbar spine conditions despite NICE LBP guidelines (2016)

recommending against imaging in non-specialist settings. This highlights the need for more powerful methods to 'nudge' clinician beliefs and behaviours (Sajid et al, 2021). Opioid prescription is another area of variation, with recent NICE guidance (2021) advising not to initiate opioid prescription for the management of chronic primary pain. In a recent feasibility trial investigating risk stratification in MSK patients seen in primary care (Hill et al, 2020), intervention practices were found to prescribe less opioids and more over-the-counter medications and anti-inflammatories, provide more self-management information and make earlier referrals to physiotherapy services than control practices. This study shows how risk stratification tools/decision aids can help to facilitate guideline concordant behaviour and could help reduce variation in practice. There is an urgent need however for more research in this area to derive and understand normative values and optimal target levels for key MSK metrics, and to collect data at scale on both process, system, and patient reported outcomes to better evaluate quality and value within MSK care pathways. This will also allow for transformation in MSK services to be captured, monitored, and evaluated to show-case exemplar services and new innovations.

Benchmarking in clinical practice involves comparing and sharing best practice (Siemens et al, 2017). Services can be compared using an agreed set of metrics, indicators, and standards (Smith et al, 2008). An example of successful MSK benchmarking within the UK, includes the NHS England led National Patient Reported Outcomes (NPROMS) Programme for hip and knee arthroplasty (NHS England, 2019), with improvement methodology supported by the 'Getting it Right First Time' (GIRFT) Programme in Orthopaedics (GIRFT, 2020). The availability of data at a surgeon, unit and trust level has equipped clinicians and managers to make informed decisions to improve the quality of patient care, including: improvement in the quality of implants used, reduction in low volume operating, and reduction in surgical site infections (GIRFT, 2020). The UK National Early Inflammatory Arthritis Audit (NEIAA) funded by the Healthcare Quality Improvement Partnership (HQIP) is another example of a national MSK audit that has successfully benchmarked provider metrics and driven forward tangible improvements in the quality of care,

including earlier referral and access to specialist rheumatology services for suspected inflammatory arthritis patients (BSR, 2021). However, whilst these existing benchmarking programmes include orthopaedic surgical patients and rheumatology data, there are no equivalent programmes in the UK benchmarking care for the vast majority of MSK patients who have non-surgical and non-inflammatory problems. For example, there are around 430,000 people with rheumatoid arthritis in the UK, 222,000 have ankylosing spondylitis and 1.6 million have gout (VA, 2019), and by contrast there are 10 million people with persistent back pain, 8.75million people with osteoarthritis, and up to 2.8 million with fibromyalgia (VA, 2019). These prevalence figures give some indication of how much more impactful these types of national audit programmes could be if expanded across the higher prevalence MSK conditions commonly managed in community/primary care.

In the wake of COVID-19 and the restoration of MSK services, it is clear that more needs to be done to define measurable indicators for community and primary care MSK services, to develop methods to identify variation in performance, to highlight and address health inequalities across MSK healthcare settings, and to deliver evidence-informed, personalised, high-quality integrated healthcare of value to all (ARMA, 2021, NHS England, 2020a, NHS Futures, 2021). Priorities set out as part of the National NHS England Best MSK Health Collaborative Programme specifically include; improving MSK data collection in community and primary care settings with use of a standardised dataset to include PROMs and PREMs, optimising appropriate use of digital resources, and having the ability to identify what works well to share with others (NHS Futures, 2021). These priorities require clear and evidence-based recommendations to ensure that they can be successfully taken forwards by MSK services. Alongside policy supporting standardisation and improved capture and reporting of quality data in MSK, is the development of evidence based MSK Service Standards (CSP, 2021a) aiming to support the delivery and development of high quality MSK physiotherapy services.

The Best MSK priorities reinforce the need to focus on MSK benchmarking, and the need to develop a national data collection in community and primary care as a national priority, to lay the foundation for effective and measurable quality improvement programmes and the sharing of best practice exemplars, and the ability for services to measure themselves against evidence informed quality standards (CSP, 2021a) as is demonstrated in the NEIAA programme (BSR, 2021).

This narrative review looked to summarise evidence from previous chapters around benchmarking in MSK services with a focus to primary and community care and to make recommendations on what metrics should be included in an MSK core outcome set, and how this work could be taken forwards into a national evaluation/audit.

Aims

1. To provide a summary of which metrics/indicators should be collected across MSK services in community and primary care as part of a minimum dataset summarising findings from previous chapters and emerging national policy.

2. To provide a summary of methods that can be used to compare these metrics including; case-mix adjustment of PROM data, deriving normative values, and use of funnel plots to identify outliers.

3. To explore how these metrics can be feasibly collected in routine clinical practice using examples from the UK.

This chapter aimed to summarise the recommendations from previous chapters (from a collection of papers focused on developing methods around benchmarking MSK services (Burgess et al, 2019, Burgess et al, 2020a, Burgess et al, 2020b, Burgess et al 2021a, Burgess et al 2021b, **Chapters 2-6**).

Which MSK indicators and metrics should we measure?

Defining the MSK Population

18.8 million people in the UK suffered with a musculoskeletal (MSK) disorder in 2017 (VA, 2019). The most prevalent MSK pain presentations in primary care include back, neck, shoulder, knee and multisite pain, with MSK consultations making up around a fifth of all primary care consultations in the UK (Jordon et al, 2010). The scope of this recommendation is outlined in **Table 7-1** below (taken from Burgess et al (2021a) **Chapter 5**).

Table 7-1: Scope of MSK Core Outcome Set (COS)

Standard Number	Domain	Methodology (COS-STAD)	Detail
1	Scope	The research or practice setting(s) in which the COS is to be applied	UK Community and Primary Care setting for use in routine clinical practice
2	Scope	The health condition(s) covered by the COS	All MSK conditions (focus on non-inflammatory, non-surgical)
3	Scope	The population(s) covered by the COS	All adult (18 years or over) MSK patients attending for routine MSK care
4	Scope	The intervention(s) covered by the COS	All interventions (focus on non-surgical)

This Table follows recommendations for reporting by the COS-STAD study team (Kirkham et al, 2017)

Further work currently underway in the Multi-level Integrated Data for Musculoskeletal Health Intelligence and Actions (MIDAS) project[3] (led by Professor George Peat (Keele University) and funded by the Nuffield Foundation and VA) looking at multi-level integrated MSK data with a focus on population health will help further define 'ontologies', this will include defining the population included within a future MSK audit by providing pre-defined SnomedCT code-lists that could then be used to identify appropriate patients and extract relevant MSK data from clinical systems.

Demographics:

[3] https://www.keele.ac.uk/midas/

A number of patient demographics/characteristics are seen as essential metrics for comparing baseline similarities in datasets and for identifying differences for certain subgroups of patients with regards to treatment effect. These are also often used to adjust scientific data for the purposes of comparison and can include descriptors such as; age, gender, ethnicity, education, and socioeconomic status. Demographics can also be used to monitor and actively address any barriers to service provision to ensure equitable access and use of NHS services (NHSE, 2020b). Within recent NHS policy/guidance in MSK community and primary care, characteristics of age, sex, and ethnicity are encouraged as a minimum (NHSE, 2020a). A recent consensus survey conducted by our research team aimed at gaining consensus on essential metrics to form part of a core outcome set in MSK. The findings identified that age and sex met strong consensus for inclusion, and ethnicity had moderate consensus (Burgess et al, 2021a, **Chapter 5**). In light of health policy and the importance of identifying and addressing inequalities in healthcare outcomes particularly in the wake of the COVID-19 pandemic, we recommend the use of age, sex, and ethnicity as a minimum to report within an MSK evaluation and benchmarked report (see **Table 7-2**). This allows for outcomes to be compared across these subgroups, potentially highlighting inequity in differing patient populations. These factors are normally readily available within electronic patient records (EPRs) and therefore should be straightforward to extract for evaluation purposes. Another factor found to be highly predictive of patient outcome is health literacy (Burgess et al 2021b, **Chapter 6**), which may also be a useful addition when evaluating MSK population health.

Clinical Factors:

Clinical factors within an MSK dataset provide an overview of clinical complexity (case-mix) of patients treated, allowing for descriptive analysis and case-mix adjustment. Factors such as; pain site, comorbidities, duration of symptoms, surgical history, previous pain episodes, disability, and previous treatment can be included (Burgess et al, 2019, **Chapter 2**). Consensus work undertaken

in 2021 showed strong consensus for the inclusion of pain site, comorbidities, and symptom duration in an MSK core outcome set, with moderate consensus for the inclusion of previous surgery, disability, and previous physiotherapy (Burgess et al, 2021a, **Chapter 5**). Patients were keen to ensure that mental wellbeing was also taken into consideration (Burgess et al 2021a, **Chapter 5**). Many of these factors are routinely collected within EPRs as part of referral, triage, and patient assessment, this data can also be collected prior to clinical visits through paper based or online patient surveys or referral forms.

In line with consensus findings, we recommend as a minimum the inclusion of pain site, comorbidities, and duration of symptoms (Burgess et al, 2021a, **Chapter 5**), with the addition of previous surgery and previous pain episodes for the purpose of case-mix adjustment (Burgess et al, 2021b, **Chapter 6**) (see **Table 7-2**).

Risk Stratification

Another mechanism for describing and analysing complexity of MSK patients in community and primary care is through the use of specific risk stratification tools. The NICE guidance for low back pain (LBP) and sciatica (NICE, 2016) recommends the use of risk stratification tools such as STarT Back (Hill et al, 2011) for each new episode of LBP to inform shared decision making and management planning. An additional risk stratification tool called STarT MSK has more recently been developed by the Keele research team extending the STarT Back to include the other most common MSK pain presentations (Campbell et al, 2016, Hill et al, 2020), making it more implementable across an MSK pathway of care. The use of risk stratification tools such as STarT Back or STarT MSK reached only moderate consensus for inclusion in a core outcome set (Burgess et al, 2021a, **Chapter 5**). We therefore recommend that risk stratification is added as an optional metric to be used for stratifying risk and guiding treatment planning at the appropriate point in an MSK pathway of care supporting best practice guidance (see **Table 7-3**).

Employment Factors:

There has been a drive over the last 10 years to increase work participation for people living with MSK disorders due to the potential health and economic benefits of remaining in employment (Wilkie et al, 2012, 2020). The FCP National Evaluation (CSP, 2020, Stynes et al, 2021) showed that 54% of patients reported less impact of their MSK condition on work at 3 months following their FCP visit, but only 29% of employed patients received work-based (vocational) advice. This shows that a greater focus is needed for this area of practice. Consensus work (Burgess et al 2021a, **Chapter 5**) identified strong support from clinicians to include metrics around employment status, work absence and work absence duration within a core outcome set. Patient feedback agreed with inclusion of work-based questions but suggested modification to the questions included to provide more response options/categories including unpaid work. Taking this feedback together with best evidence and the feasibility of including work based questions into a routine dataset we recommend the inclusion of 4 modified questions from the Work Productivity and Activity Impairment Questionnaire (WPAI) (see **Table 7-2**). The WPAI has been shown to be quick to fill out, valid in measuring absenteeism (time off work) and presenteeism (productivity loss), and is free to use (Reilly et al, 1993, Noben et al, 2014, Wilkie et al, 2020).

Patient Reported Outcome Measures (PROMs):

Patient Reported Outcome Measures (PROMs) capture patient's own opinions on their health and capture the impact of their health condition and treatment on their life using short self-completed items (Kyte et al, 2015, Hill et al, 2016).

A systematic review conducted in 2018 looked to identify PROMs utilised by UK Advanced Practice Physiotherapists in MSK (Fennelly et al, 2018). This review found that the most frequently used PROMs included patient satisfaction, quality of life (QoL), functional status, and pain, and less frequently global improvement, mental well-being, work ability, and healthcare

utilisation/costs (Fennelly et al, 2018). The EQ5D and SF36 were identified as the most commonly reported QoL tools, visual analogue scales (VAS) were most used to capture pain, but there was uncertainty about how to capture functional status. Findings from a recent consensus survey in 2021 (Burgess et al 2021a, **Chapter 5**) supported the importance of patient reported experience measure (PREM) domains but there was less certainty (no strong consensus) from clinicians around use of PROMs, with the strongest consensus (at the moderate level) reached for use of the Musculoskeletal Health Questionnaire (MSK-HQ). The EQ5D (Euroqol, 2019) did not reach strong or moderate consensus levels in this study and therefore was removed from the potential list of metrics for inclusion in a standardised core outcome set (Burgess et al, 2021a, **Chapter 5**). Pain intensity using a numeric pain rating scale (NPRS) (Downie et al, 1978) also met moderate consensus (Burgess et al, 2021a, **Chapter 5**), and a small number of respondents also reported using the Patient Specific Functional Scale (PSFS) (Stratford et al, 1995). Previous recommendations within the Versus Arthritis Musculoskeletal Indicator Set (ARUK, 2016) recommend the use of the MSK-HQ or the EQ5D to measure MSK health utility outcomes (Indicator 20), but this is still not standard practice across MSK services with 17% of healthcare professionals in the consensus survey by Burgess et al (2021a, **Chapter 5**) reporting that they are not capturing any PROMs in routine care.

The MSK-HQ was co-produced with patients and designed for use across a clinical pathway of care as a means to capture MSK health status. It has been shown to be valid, reliable and sensitive to change with strong convergent validity against reference standards including the commonly used EQ5D5L and MSK disease specific tools (Oxford Hip, Knee and Shoulder scores) (Hill et al, 2016, Price et al, 2019a). Price et al (2019a) showed that the MSK-HQ responsiveness was superior to the EQ5D5L across all subgroups and that it showed strong correlation with the disease specific Oxford hip, knee and shoulder measures. The MSK-HQ is therefore recommended as the generic PROM of choice to use across an MSK pathway of care (see **Table 7-2**). The minimum clinical important difference (MCID) for this measure was shown to be 5.5 (Price

et al, 2019a). The percentage of patients improving by more than 6 points on the MSK-HQ was evaluated as part of the national FCP evaluation with pre-defined success criteria for treatment effectiveness being met if this figure surpassed 51% (CSP, 2020, Stynes, 2021).

Patient Reported Experience Measures (PREMs):

Patient experience alongside clinical effectiveness and patient safety is one of the three key quality pillars in healthcare and its measurement is often mandatory (Doyle et al, 2013). For example, there is a minimum requirement for NHS services to use the Friends and Family Test (FFT) (NHS England, 2019). Evidence supports the notion that better patient care experiences are associated with improved treatment compliance and better clinical outcomes (Anhang Price et al, 2014, Smith and Choma, 2017). Despite the importance of PREMs as quality metrics however there remains a lack of standardisation for measuring 'experience' or 'satisfaction' with care in MSK practice (Roberts, 2012).

Consensus on key MSK metrics from MSK healthcare professionals showed that PREMs are seen as a priority to collect in clinical practice, nine out of ten listed PREM domains in the consensus study met strong consensus making up by far the largest group of metrics to meet strong consensus (Burgess et al, 2021a, **Chapter 5**). Listed PREM domains included; being shown care and respect, being understood and valued, the FFT, confidence in clinical competence, involved in shared decision-making, given sufficient time, patient perceived overall improvement, improved understanding of health condition, improved confidence to manage yourself, and timeliness and convenience of consultations, with the last domain the only one not to meet strong consensus. Patients also supported collection of PREM domains listed and additionally suggested inclusion of questions around access to MSK care (Burgess et al, 2021a, **Chapter 5**). Access and making appointments was also an area of focus for the updated NHS funded GP Patient Survey (GPPS) (NHS England and Ipsos MORI, 2018). Interestingly although healthcare professionals strongly agreed on inclusion of PREM domains in a core MSK outcome set, 25% of healthcare

professionals reported they are not currently collecting these in routine care (Burgess et al, 2021a, **Chapter 5**) which reflects the difficulty with embedding these tools successfully into clinical practice in the absence of a funded programme. A national group of MSK Stakeholders in the UK is currently developing a national PREM to be used in MSK services taking all of these findings into consideration (see **Appendix 7-1**). This group is working closely with NHS England and is also incorporating work from the GP Patient Survey (GPPS) (NHS England and Ipsos MORI, 2018), the FFT developed by NHS England (NHS England, 2019), and findings from the consensus study conducted by Burgess et al (2021a, **Chapter 5**).

Health utilisation and cost indicators:

Porter describes value-based care as the health outcomes achieved per monetary unit spent (Porter, 2010). Healthcare transformation designed to improve access and quality of care whilst containing cost is an international priority discussed by Speerin et al (2020) in their review of models of care for MSK conditions supporting a value-based care approach. One mechanism for this approach in MSK has been through allowing advanced physiotherapy practitioners to work to their maximum scope as seen within the UK with the First Contact Practitioner (FCP) in primary care agenda (Speerin et al, 2020, CSP, 2020). In order to evaluate costs and use of healthcare resources across MSK pathways of care and emerging models of practice, routine collection and reporting of key MSK cost drivers is needed. A systematic review of key MSK cost drivers in 2020 (Burgess, 2020, **Chapter 4**) found that GP visits, outpatient visits, and physiotherapy visits were by far the highest mean costs across included economic analyses, followed by prescription medications and hospital admissions. These five cost drivers alone captured over 70% of the costs in the majority of included studies and could provide a feasible method for estimating costs in MSK services and across a pathway of MSK care. Collection of MSK cost data however is seen as less of a priority by healthcare professionals with no cost metrics/economic factors meeting strong consensus for inclusion in an MSK core outcome set (Burgess et al, 2021a, **Chapter 5**). For

this reason, cost indicators were added to the 'optional' metrics in our recommended list (see **Table 7-3**).

Key Recommendations

Table 7-2 outlines metrics for inclusion in a minimum dataset, with **Table 7-3** outlining additional optional metrics which would ideally be collected alongside those in **Table 7-2** where systems are in place for seamless, integrated, data capture. The core set includes: demographics; **age, sex, ethnicity,** clinical factors; **pain site, comorbidities, duration of symptoms, previous surgery, previous pain episodes**, PROM; **MSK-HQ**, PREM; **National MSK PREM** (see **Appendix 7-1** for detail) which includes domains informed by findings in **Chapter 5** of; **access/waiting times, shared decision making, confidence in clinicians, treated with care and respect, time and understanding, care planning, information giving, convenience/timeliness, needs met and ideas to improve our service**[4].

[4] Link to newly developed National PREM (unpublished (also see **Appendix 7-1** for Abstract submitted to the Centre for Advancing Practice 2021)): https://forms.office.com/Pages/ResponsePage.aspx?id=V2N9w4vIa0K2gN-BZqhu131I5H8TIg5PvDFb_xI7-eZUMko4NkxQRU5GMkY0NENSTjBGQUw3NjY0WS4u

Table 7-2 Keele MSK Core Outcome Set: Mandatory Variables

Variable Name	Response Options	Capture Point	Capture From	Purpose	Evidence to support inclusion
Mandatory Variables					
Demographics					
Age	Continuous numeric	Baseline	EPR	Descriptive	Burgess et al 2021a, NHSE 2020
Sex at birth	Binary (male/female)	Baseline	EPR	Descriptive	Burgess et al 2021a, NHSE 2020
Ethnicity	Categorical (5 options)	Baseline	EPR	Descriptive	NHSE 2020
Clinical Factors					
Pain Site	Categorical (11 options)	Baseline	EPR/Survey	Descriptive	Burgess et al 2019, 2021a
Comorbidities	Categorical (12 options)	Baseline	EPR/Survey	Case-mix complexity	Burgess et al 2019, 2021a, 2021b
Duration of Symptoms	Categorical (5 options)	Baseline	Survey	Case-mix complexity	Burgess et al 2019, 2021a, 2021b
Previous Surgery	Binary (yes/no)	Baseline	Survey	Case-mix complexity	Burgess et al 2019, 2021b
Previous Pain Episodes	Binary (yes/no)	Baseline	Survey	Case-mix complexity	Burgess et al 2021b
Employment					
Work Productivity and Activity Impairment Questionnaire (WPAI)	Questionnaire (4 questions (modified))	Baseline & 3 months	Survey	Work PROM Case-mix complexity	Burgess et al 2019, 2021a, 2021b, Reilly et al, 1993, Wilkie et al, 2020
Functional Status					
MSK-HQ (MSK Health Status)	Questionnaire (15 questions)	Baseline & 3 months	Survey	MSK PROM Case-mix complexity	Burgess et al 2021a, Hill et al 2016, Price et al 2019a
Patient Reported Experience					
National MSK PREM*	Questionnaire (11 questions)	3 months	Survey	PREM	Burgess et al 2021a GPPS 2018, NHSE 2019, 2020

EPR; Electronic Patient Record, GPPS; GP Patient Survey (NHS England and Ipsos MORI, 2018), NHSE; NHS England, PREM; Patient Reported Experience Measure, PROM; Patient Reported Outcome Measure.

* National MSK PREM currently unpublished (see **Appendix 7-1** for Abstract presented at the Centre for Advancing Practice Conference 2021)

Table 7-3 Keele MSK Core Outcome Set: Optional Variables

Variable Name	Response Options	Capture Point	Capture From	Purpose	Evidence to support inclusion
Optional Variables					
Demographics					
Health Literacy	Categorical (5 options)	Baseline	Survey	Descriptive	Burgess et al 2021b
Clinical Factors					
Previous Physiotherapy	Binary (yes/no)	Baseline	Survey	Descriptive	Burgess et al 2021a
Functional Status					
STarT Back/MSK (Risk Status)	Questionnaire (10 questions)	Baseline	Survey	Risk stratification	NICE Guidelines 2016, Burgess et al 2021a, Hill et al 2011, Campbell et al, 2016
Numeric Pain Rating Scale	Numeric VAS (0-10)	Baseline & 3 months	Survey	PROM	Burgess et al 2021a, Downie et al 1978
Patient Specific Functional Scale	Questionnaire	Baseline & 3 months	Survey	PROM	Burgess et al 2021a, Stratford 1995
Economic Factors					
Healthcare Utilisation	Free text numeric	3 months	Survey/EPR	Cost indicator	Burgess et al 2020, 2021a
Investigations and Treatments	Free text numeric	3 months	Survey/EPR	Cost indicator	Burgess et al 2020, 2021a
Inpatient Stays	Free text numeric	3 months	Survey/EPR	Cost indicator	Burgess et al 2020, 2021a
Prescribed Medication	Binary (yes/no)	3 months	Survey/EPR	Cost indicator	Burgess et al 2020, 2021a

EPR; Electronic Patient Record, VAS; Visual Analogue Scale.

What methods should we use to compare MSK service performance?

Case-mix adjustment: Case-mix adjustment is a statistical process that aims to account for differences in the mix of patient attributes/characteristics across patient groups (e.g., those treated by different healthcare providers) in order to make fair comparisons of the relative effectiveness (outcome) of care provided (Iezzoni, 2009). Case-mix adjustment allows for outcomes to be compared on a like-for-like basis across providers (Coles, 2010) to ensure that those seeing the most complex patients are not disadvantaged. Adjusting data across providers allows for meaningful comparisons to be made with identification of positive and negative outliers that can then be used as exemplars to inform quality improvement. Our research team previously identified two existing MSK case-mix models (Burgess et al, 2019, **Chapter 2**), one for use in US community/primary care clinic settings (FOTO model, Deutscher et al (2018)), and one developed for use in a secondary care surgical population in the UK (UK National PROMs Model, Coles 2010, DoH 2012). These existing (modified) MSK models were later externally validated for use in a UK community/primary care population (Burgess et al, 2021b, **Chapter 6**) and were found to be highly valid in this setting predicting 44% and 41% of the variance in primary outcome respectively (MSK-HQ PROM score). Key variables within the slightly more predictive FOTO model were: baseline PROM score, previous pain episodes, comorbidities, current paid employment, previous surgery, physical activity, and duration of symptoms. These factors provide a feasible list of metrics which if included as part of a minimum dataset would allow for complexity of caseloads to be taken into consideration enabling fair and transparent comparisons of performance/outcomes.

The method used by NHS England to adjust, compare, and identify outliers within the NPROMS hip and knee arthroplasty data is described in brief below to provide an example of robust methods for benchmarking in routine MSK care that could be adapted and utilised to adjust community and primary care MSK data.

The NPROMS benchmarking method: Case-mix models were developed and internally validated separately for each of the PROM measures (Oxford Hip Score, Oxford Knee Score, EQ5D, EQ5D Visual Analogue Scale (VAS)) collected within the National PROMs hip and knee arthroplasty audit programme (Coles et al, 2010). The case-mix adjustment process has three key stages; estimation of the impact of case-mix variables, generation of patient-level predicted scores, aggregation to organisation level and case-mix adjustment (NHS England, 2013). Funnel plots are then used to compare providers' performance[5]. Providers' adjusted average health gain scores are compared to the mean national figure, and control limits at two and three standard deviations from this are used to compare the health gain status of providers, and to identify positive and negative outliers (NHS England, 2013). These charts are used widely across NHS systems including the Model Health System[6] for benchmarking and help improve visualisation of data (NHS England and NHS Improvement, 2021).

Recommendation for Community/Primary Care benchmarking: In order to replicate the NPROMS method across community/primary care services, firstly; metrics and data capture would need to be standardised and successfully adopted/implemented across providers (use of data dictionary for integration into clinical systems linking through to the integrated care record (NHS Digital, 2021), secondly; data would need to be uploaded to a central repository for advanced analytics including case-mix adjustment and outlier identification (this would need to be funded/commissioned in order for this to be a continuous process of data upload-advanced analytics-dashboards, and could be housed within existing NHS systems such as the Community Services Dataset (CSDS) and/or Model Health System (NHS England and NHS Improvement, 2021), and thirdly; benchmarked reports would need to be developed and disseminated to form the basis for quality improvement (QI) (benchmarks can be derived from within a dataset using

[5] See; https://www.england.nhs.uk/wp-content/uploads/2018/08/proms-guide-aug-18-v3.pdf for explanation of funnel plots and how to identify outliers.
[6] See; https://feedback.model.nhs.uk/knowledgebase/articles/1979268-funnel-plots

the statistical mean of the data, with positive outliers used as exemplars of optimal models of care to inform QI initiatives (DoH, 2012)). Reports would need to be co-developed with stakeholders including patient partners to ensure that data was made meaningful at all levels to maximise its impact and usability. Quality improvement methodology such as that used by the 'Getting it Right First Time' (GIRFT) initiative (GIRFT, 2021) could then be adopted to support implementation of recommendations, along with the use of specific MSK quality standards such as those developed and due for release imminently by the CSP (CSP, 2021a).

What methods should we use to collect MSK metrics in routine MSK care?

Examples from the UK

Collecting and reporting metrics in routine care is challenging and made more complex by the differing electronic medical record systems used across the UK healthcare sector and the difficulties faced to fully integrate these systems to allow for optimal extraction and reporting of MSK data at a local, system and national level. Four MSK/Pain-focused large national data collections have been explored below to highlight methods of collection and response rates within existing or previously funded audits and evaluations in MSK.

NPROMS: Since April 2009 all providers of NHS-funded unilateral hip and knee replacements in England have been required to collect and report PROMs (NHS Digital, 2017). A mix of online and paper questionnaires are collected pre and post procedure and are linked routinely with the Hospital Episode Statistic (HES) database for further episode level information. In April 2019 to March 2020, 76,401 hip procedures were carried out. 67,492 questionnaires were returned pre-operatively (88.3%) and 41,854 (54.8%) post-procedure. There were 35,937 paired datasets linked to an episode of care giving 47% of patients in total with complete and matched data (35,937/76,401) (NHS Digital, 2020). For knees 90,309 procedures were undertaken across England with 79,803 (88.4%) questionnaires returned pre-procedure and 48,116 (53.3%) post-procedure. There were 41,319 paired/matched datasets linked to an episode of care (45.8%)

(41,319/90,309) (NHS Digital, 2020). This means out of all hip and knee procedures undertaken in England, 47% and 46% respectively had pre- and post-operative matched data (see **Table 7-4**). Pre-operative completion of PROM questionnaires is linked to best practice tariff (BPT) thus incentivising providers to comply with the national audit programme (Gutacker et al, 2015), this is reflected in the high baseline response rates.

NEIAA: The National Inflammatory Arthritis Audit (NEIAA) was setup to improve the quality of care for people living with inflammatory arthritis. In the second year of the audit (2019-2020) 118/137 trusts submitted data (BSR, 2021). A total of 13,578 patients with suspected inflammatory arthritis were seen in rheumatology services in England and Wales in 2019/20. A large proportion of the NEIAA data is entered by clinicians/services through an online portal with good uptake of clinician entered data reported, e.g., 12653/13,578 patients (93%) had a recorded diagnosis. All patients eligible for Early Inflammatory Arthritis (EIA) follow up were also invited to fill out PROMs at baseline, 3 and 12 months, patients completed this online or via paper in clinic. A digital partner (Netsolving) was commissioned to provide an integrated solution to online PROM data collection. Across the 2 years of data capture, 41% (4996/12,185) of eligible EIA patients completed PROM data at baseline, 20% (2482/12,185) at 3 months (BSR, 2021).

FCP Evaluation: The Chartered Society of Physiotherapy funded a national evaluation of FCP services in 2019-2020 (CSP, 2020, Stynes et al, 2021). The purpose of this evaluation was to assess the FCP model of MSK primary care against pre-defined quality standards. Forty services participated in the evaluation with 2825 patients registered and 680 (24%) patients with complete data at baseline for inclusion. Data collection was in the form of an electronic patient survey questionnaire with software commissioned from an industry partner called PRO-MAPP. Follow-up rates at 1, 2, and 3 months, were 63% (n=430), 62% (n=419) and 54% (n=370) respectively. Barriers to clinicians registering patients for the study included time constraints, language barriers and the pressures of a new role.

National Pain Audit: Price et al (2019b) reported an evaluation of specialist pain services with data collected as part of a UK Department of Health (DoH) funded national pain audit from 2010-2014. The focus of this paper was on feasibility with regards to data completion and overall response rates. Of those patients eligible to participate (49,460), 19% (9558 patients) were recruited. There was 92% item completion rate for survey data which was collected on paper. 46% (4414) of included patients completed data at 3 months, and 19% (1799) at 12 months. Barriers reported included; poor coding and classification in electronic systems, fear of scrutiny and increased workload from providers, and lack of ongoing funding being a bigger factor beyond the remit of the funded audit. Recommendations for increasing response rates included public feedback, clear explanation of goals and dedicated staff to oversee PROM collection.

Table 7-4: Completion Rates for PROM data in National Audits/Evaluations in MSK

Audit/Evaluation Reference Funder	Total Eligible Population	Complete data at baseline (%)	Complete (matched) data at follow up (% of total eligible, % of baseline)	Follow up time-point	Method of collection (online provider)	Incentivised (Yes/No)	Benchmarked Report
NPROMs Hip (NHS Digital, 2020) Funded by NHSE	76,401	67,492 (88%)	35,937 (47%, 53%)	6 months	Paper or online survey & HES matching	Yes (linked to BPT)	Yes with outlier identification
NPROMs Knee (NHS Digital, 2020) Funded by NHSE	90,309	79,803 (88%)	41,319 (46%, 52%)	6 months	Paper or online survey & HES matching	Yes (linked to BPT)	Yes with outlier identification
NEIAA (BSR, 2021) Funded by HQIP	12,185	4996 (41%)	2,482 (20%, 50%)	3 months	Paper or Online survey and clinician web tool (Net Solving)	Not for patient reported data	Yes but no outlier identification for PROM data
National FCP Evaluation (CSP, 2020) Funded by CSP	2,825	680 (24%)	370 (13%, 54%)	3 months	Online survey (PRO-MAPP)	No	No
National Pain Audit (Price et al, 2019) Funded by DoH	49,460	9558 (19%)	4414 (9%, 46%)	3 months	Paper survey and clinician web tool (Dr Foster)	No	No

BPT; Best Practice Tariff, BSR; British Society of Rheumatology, CSP; Chartered Society of Physiotherapy, DoH; Department of Health, HES; Hospital Episode Statistics, NHSE; NHS England, NPROMs; National Patient Reported Outcome Measures Programme.

Recommendation for data capture: It is clear from published evaluations, audits and research studies that large scale collection of patient questionnaires/surveys and MSK metrics is highly challenging. Optimal uptake is shown in areas where the evaluation/audit process is both funded and incentivised with clear reporting of data. This demonstrates that more support is needed for clinical services if uptake of a core outcome set within MSK community/primary care is to be successful. Recommendations for successful implementation include:

- Prioritisation and investment from policy holders/funders of care; to invest in appropriate software, analytics, training and dissemination with consideration to financial incentives to optimise uptake.

- Development of digital infrastructure and integration; linking EPR systems and digital applications (supported by robust governance), developing intra-operability, allowing for more streamlined approaches to data collection, sharing, and reporting, to make routine data collection as part of a local or national evaluation/audit more feasible.

- Setting realistic goals: For services developing electronic systems and looking to adopt recommendations around MSK metric capture, realistic initial targets need to be set with stakeholders with regards to response rates at intake and specified follow up, for example aiming for 40% completion at baseline and 20% at 3 month follow up in line with other national programmes such as NEIAA (BSR, 2021).

- Co-development with patients and clinicians; for successful uptake locally, systems need to take into consideration local population needs such as health and digital literacy as this may impact on uptake of PROM capture (we know that 22% of the population do not have the digital skills needed for everyday life (NHS Digital, 2021b)). Clinicians also need to feel engaged with data collection to allow for effective utilisation of patient survey

data and engagement with inputting data correctly (clinical templates/coding/online portals). Clinical PROM champions may aid PROM capture locally.

Conclusion

It is clear that there is a need for a standardised approach to collecting metrics for common MSK conditions seen in primary/community care. Answers to three important questions have been explored in this Chapter, and a summary of findings/recommendations are provided:

1. **Which MSK indicators and metrics should we measure?**

   A core minimum dataset for MSK services in this setting should include: demographics; **age, sex, ethnicity,** clinical factors; **pain site, comorbidities, duration of symptoms, previous surgery, previous pain episodes**, PROM; **MSK-HQ,** PREM; **National MSK PREM** which will include domains of; access/waiting times, shared decision making, confidence in clinicians, treated with care and respect, time and understanding, care planning, information giving, convenience/timeliness, needs met and ideas to improve our service.

2. **What Methods should we use to compare MSK service performance?**

   1. Implementation/collection of the standardised minimum core outcome set (**Table 7-2**) plus or minus optional extras (**Table 7-3**) across multiple MSK services; 2. Data uploaded to a central repository for advanced analytics including case-mix adjustment and outlier identification; 3. Benchmarked reports developed and disseminated to form the basis for quality improvement (QI), outcomes reviewed against national quality standards (CSP, 2021), and implementation of QI following GIRFT methodology (GIRFT, 2021).

3. **What methods should we use to collect MSK metrics in routine MSK care?**

1. Prioritisation/investment/leadership from policy holders/funders of care: Capture of MSK data is highlighted as an urgent priority in UK national policy, with the Best MSK Health Programme outlining the need for standardised quality data in community/primary care (NHSE, 2020a, NHS Futures, 2021). Funding and leadership will be the key to successful implementation of a core outcome set which needs; 2. Development of digital infrastructure/integration/governance to allow for widespread capture of MSK metrics including PROMs and PREMs, ease of matching data fields across multiple electronic systems at different time-points, an agreed back end data repository to house big data, and automated reporting for routine feedback; 3. Engagement; Engagement of the MSK community including clinicians, patients, and additional MSK stakeholders, with feedback captured to overcome barriers, and setting of realistic goals/targets with stakeholders with regards to response rates giving consideration to resources provided and financial incentives to allow for appropriate expectation.

Thesis Summary and next steps

Successful implementation of widespread standardised data capture would allow for development of an MSK learning health system, whereby we not only see research informing clinical practice, but we also see large routine data collections informing research priorities. By using robust and transparent processes to inform us where the greatest population health needs/inequities/variation in MSK outcomes are, and where we see optimum value in MSK care pathways, research priorities could be generated, and quality improvement methodologies/transformation of services undertaken with further analysis explored to maximise benefit to all.

Standardising, collecting, extracting, analysing, comparing, and reporting MSK data for the purpose of evaluating quality/performance and improving patient care is highly challenging.

More support and guidance is needed in this area if clinical services are to successfully adopt recommendations from this thesis into practice. Some of the key challenges include leadership, funding, digital infrastructure and integration, and analytical capabilities. These challenges should provide part of the focus for the UK BEST MSK Health Collaborative Programme (NHSE, 2020a, NHS Futures, 2021) with a longer-term plan made for how to enable and embed effective data capture and analytics to evidence and inform best MSK care for the future.

Priorities for future research identified within this research thesis include; the need to test methods for routine big data collection in MSK clinical practice across community and primary care services (developing digital infrastructure and advanced analytics), the need to engage the MSK community with the need for enhanced MSK data collection and identify barriers and enablers to successful implementation of a national audit, and the need to measure MSK services against clear quality standards to identify variation and best practice exemplars. These priorities have been used to develop a post-doctoral research proposal which is summarised in brief in **Figure 7-1**.

The research proposal (**Figure 7-1**) has been underpinned by thesis research outputs and is a clear, logical, and timely next step to the programme of work conducted to date. This project aims to address the gap in data provision by developing and testing a standardised approach to collecting and reporting community and primary care MSK data to facilitate quality improvement. Identifying variation in quality, practice, value and efficiency is fundamental across the health system, so rather than adopt an approach specific to condition we felt it would be more equitable and powerful to adopt a quality approach focused to including all MSK conditions and with a focus to patients presenting in primary care MSK clinics through First Contact Practitioners (FCPs) and through community MSK services in the first instance, to better identify gaps and inequities in MSK provision and performance.

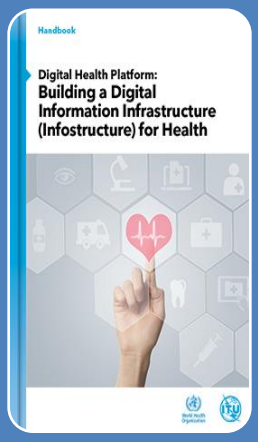
Specific performance areas already identified as problematic in MSK services include; delays in access to care, failure to provide self-management information including vocational advice and home exercises, overuse of imaging, overuse of opioids, overuse of surgery, and overuse of emergency services (Lin et al, 2020, Hill et al, 2020, CSP, 2020, Stynes et al, 2021). These areas will be evaluated alongside system changes such as the adoption of the MSK FCP roles in primary care and the move to virtual and telephone consultations adopted in response to the COVID-19 pandemic (NHS England, 2020a). Variation in practice will be identified alongside quality metrics including patient reported health gain, return to work and patient satisfaction, and additionally key cost indicators that include healthcare visits, prescriptions, and investigations (Burgess et al, 2020b, **Chapter 4**) as outlined above, to identify optimal models of care. These metrics will be captured as part of the system, process and patient reported data. Service performance will then be measured against national quality standards (e.g., those measured in the national FCP pilot led by Keele University (Stynes et al, 2021) and those being developed by the CSP (2021a)). This work will also be aligned to the MIDAS project (led by Professor George Peat (Keele University) and funded by the Nuffield Foundation and VA) looking at multi-level integrated MSK data with a focus on population health. This will allow for results of the audit to be compared to findings from the MIDAS study that used digital, telephone supported, and paper capture, to evaluate how representative the audit participant population are of the local population of MSK patients.

From previous projects in this area (FCP National Evaluation, MIDAS), we already have a workforce that is keen to participate and ready to recruit patients into this type of study, and from the thesis outputs; a proposed case-mix model that would allow us to appropriately compare MSK health outcomes (Burgess et al 2021b, **Chapter 6**), and a standardised MSK core outcome set to help inform data capture (Burgess et al, 2021a, **Chapter 5**), and through the Keele led MIDAS project we are developing methods around data integration from multiple sources which will help inform digital infrastructure/solutions.

In summary at present there is limited national leadership but an emerging interest and focus on MSK metrics, data standards, and quality improvement through the Best MSK Health Programme led by NHS England. There is however no formal or structured prospective data capture of quality data across community and primary care MSK services including data on high prevalence MSK conditions such as LBP and osteoarthritis in the UK. This means at present we do not have the ability to continuously monitor quality in MSK community/primary care and to see the full impact of system and local level transformation. This has been highlighted as an area of interest by groups such as Pfizer with the need identified for high quality research in this area to help find solutions to better facilitate real world research in MSK conditions and chronic pain (Pfizer Ltd, 2021). The research team would now like to conduct an independent programme of research to address this lack of real-world research in MSK community/primary care and develop the methodology, infrastructure, stakeholder engagement, and capability for a sustainable national MSK audit programme, supporting national policy and the Best MSK Health Collaborative in delivering high quality care of value to all (NHS Futures, 2021).

Figure 7-1 Future Research Proposal Summary

Phase 1: Setup: Infrastructure, Governance and Engagement of Services (Months 1-12)



- **Protocol Writing and HRA/Ethical Approvals** (with patient input)
- **Infrastructure and Governance :** 1. Agree the method for each registered MSK service to collect a bespoke patient survey and clinician data collection for 2 years from multiple sites with interoperability to upload to back end data warehouse. 2. Start contract development with NHS England/NHS Digital (data controller) and appropriate third parties to provide the back- end data warehouse (e.g. Community Services Dataset (CSDS); analytics tools/query tools, map data flow, data integration/matching)
- **Engagement:** National recruitment of up to 40 MSK services across the UK (FCPs/GPs and MSK physiotherapy services)
- **Data Dictionaries and Reporting Templates :** Development of data dictionary and clinical templates with a view to automated reports/visualisations.
- **PROM Platform:** Development of PROM platform content for patient inputted data with clear process developed for how this is automated from clinical systems for primary care/FCP and community MSK clinics.
- **Site Setup:** Setup data sharing agreements with recruited MSK services and training for participating staff in clinical templates and audit infrastructure.
- **Integration:** Explore methods to link service level PROM data to patient data from other sources e.g. place-based data (link to MIDAS project findings).

Phase 2: Data Collection and Reporting (Months 12-24)

- **Data Collection:** Prospective cohort of routine data collection for initial 12-month period across 40 services (with purposive sampling from primary care (FCP) and community MSK services across the UK), with follow up data and full survey measures at baseline and at 3, 6 months, collected online through a digital patient platform and through a clinician inputted Electronic Patient Record (EPR) template.
- **Analytics:** Develop patient and provider analytics with bespoke reports including a patient summary document, provider reports, visualisations and infographics to help explain the data, working with patient, clinical/provider and other stakeholder partners.

Phase 3: Benchmarking, Quality Improvement Plans, and Developing a Roadmap (Months 24-36)

- **Data Collection:** Continued data collection as above
- **Benchmarking**: Data analysis; adjustment for case-mix, development of specific quality benchmarks for non-inflammatory MSK, identification of positive and negative outliers (use of funnel plots) and development of bespoke reports
- **Quality Improvement:** Data reviewed against national quality standards and fed back to services as part of benchmarked reports to facilitate local QI.
- **Feedback:** Semi-structured interviews with patients (n=10-20), and focus groups with clinicians (n=20-25) to gain insights about their engagement and experiences from participating in the pilot audit and their perceptions of the utility of reports and interactive analytics tools, and wider stakeholders (n=20-25) to gain insights about perceptions of the utility of reports and interactive analytics tools exploring which aspects of the audit need further optimising.
- **Development of Roadmap:** Making recommendations for a future sustainable national audit programme including financial model.

Image 1: World Health Organisation (2020), Image 2: NHS England (2021), Image 3: CSP (2021b)

8 References

Chapter 1:

Ahmed, S., Berzon, R.A., Revicki, D.A., Lenderking, W.R., Moinpour, C.M., Basch, E., Reeve, B.B. and Wu, A.W., 2012. The use of patient-reported outcomes (PRO) within comparative effectiveness research: implications for clinical practice and health care policy. *Medical care*, pp.1060-1070.

ARUK, 2016. MSK recommended indicator set for healthcare professionals. Available at: https://www.versusarthritis.org/policy/resources-for-policy-makers/for-healthcare-practitioners-and-commissioners/msk-recommended-indicator-set/msk-recommended-indicator-set-for-healthcare-professionals/ Accessed 26/10/21.

Austin, P.C., 2011. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate behavioral research*, *46*(3), pp.399-424.

Black, N., 2013. Patient reported outcome measures could help transform healthcare. *BMJ: British Medical Journal*, *346*.

Blyth, F.M., Briggs, A.M., Schneider, C.H., Hoy, D.G. and March, L.M., 2019. The global burden of musculoskeletal pain—where to from here?. *American journal of public health*, *109*(1), pp.35-40.

Briggs, A.M., Shiffman, J., Shawar, Y.R., Åkesson, K., Ali, N. and Woolf, A.D., 2020. Global health policy in the 21st century: Challenges and opportunities to arrest the global disability burden from musculoskeletal health conditions. *Best Practice & Research Clinical Rheumatology*, p.101549.

British Society of Rheumatology, 2021. National Early Inflammatory Arthritis Audit – Second Annual Report. 8 May 2019 - 7 May 2020. Healthcare Quality Improvement Partnership (HQIP). Available at: https://www.rheumatology.org.uk/practice-quality/audits/neiaa Accessed 27/10/21

Brucker, S.Y., Schumacher, C., Sohn, C., Rezai, M., Bamberg, M. and Wallwiener, D., 2008. Benchmarking the quality of breast cancer care in a nationwide voluntary system: the first five-year results (2003–2007) from Germany as a proof of concept. *BMC cancer*, *8*(1), p.358.

Childs, J.D., Harman, J.S., Rodeghero, J.R., Horn, M. and George, S.Z., 2014. Implications of practice setting on clinical outcomes and efficiency of care in the delivery of physical therapy services. *Journal of orthopaedic & sports physical therapy*, 44(12), pp.955-963.

Clement, R.C., Welander, A., Stowell, C., Cha, T.D., Chen, J.L., Davies, M., Fairbank, J.C., Foley, K.T., Gehrchen, M., Hagg, O. and Jacobs, W.C., 2015. A proposed set of metrics for standardized outcome reporting in the management of low back pain. *Acta orthopaedica*, *86*(5), pp.523-533.

Coles, J. PROMs risk adjustment methodology guide for general surgery and orthopaedic procedures. Available at: https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2013/07/proms-ris-adj-meth-sur-orth.pdf Accessed 27/10/21

Cornish, J.A., Tekkis, P.P., Tan, E., Tilney, H.S., Thompson, M.R. and Smith, J.J., 2011. The national bowel cancer audit project: the impact of organisational structure on outcome in operative bowel cancer within the United Kingdom. *Surgical oncology,* 20(2), pp.e72-e77.

Curtis, L., 2009. Unit costs of health and social care 2014. Personal Social Services Research Unit. Canterbury. Available at: https://www.pssru.ac.uk/project-pages/unit-costs/unit-costs-2009/ Accessed 27/10/21

Curtis, L., Burns, A., 2016. Unit costs of health and social care 2016. Personal Social Services Research Unit. Canterbury. Available at: https://www.pssru.ac.uk/project-pages/unit-costs/unit-costs-2016/ Accessed 27/10/21

Darzi, A., 2008. High Quality Care For All: NHS next stage review final report. Cm 7432. London: the Stationery Office. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/228836/7432.pdf Accessed 27/10/21

de Rooij, M., van Der Leeden, M., Heymans, M.W., Holla, J.F., Häkkinen, A., Lems, W.F., Roorda, L.D., Veenhof, C., Sanchez-Ramirez, D.C., de Vet, H.C. and Dekker, J., 2016. Course and predictors of pain and physical functioning in patients with hip osteoarthritis: Systematic review and meta-analysis. *Journal of rehabilitation medicine,* 48(3), pp.245-252.

Deeks, J.J., Dinnes, J., D'Amico, R., Sowden, A.J., Sakarovitch, C., Song, F., Petticrew, M. and Altman, D.G., 2003. Evaluating non-randomised intervention studies. *Health technology assessment (Winchester, England)*, *7*(27), pp.iii-173.

Department of Health, 2010. National Schedule of Reference Costs 2008-2009. London: Department of Health. Available at: https://data.gov.uk/dataset/f9b1a80c-187e-4e92-9fe1-25370291f5c0/nhs-reference-costs-2008-09 Accessed 27/10/21

Department of Health, 2011. NHS Reference costs 2009–2010. Available at: https://www.gov.uk/government/publications/nhs-reference-costs-2009-2010 Accessed 27/10/21.

Department of Health., 2012. PROMs in England; the case-mix adjustment methodology. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/216507/dh_133449.pdf Accessed 27/10/21

Devlin, N., Appleby, J., 2010. Getting the Most out of PROMs: Putting Health Outcomes at the Heart of NHS Decision Making, London: The King's Fund. Available at: https://www.kingsfund.org.uk/sites/default/files/Getting-the-most-out-of-PROMs-Nancy-Devlin-John-Appleby-Kings-Fund-March-2010.pdf Accessed 27/10/21.

Deutscher, D., Horn, S.D., Dickstein, R., Hart, D.L., Smout, R.J., Gutvirtz, M. and Ariel, I., 2009. Associations between treatment processes, patient characteristics, and outcomes in outpatient physical therapy practice. *Archives of physical medicine and rehabilitation*, 90(8), pp.1349-1363.

Deutscher, D., Werneke, M.W., Gottlieb, D., Fritz, J.M. and Resnik, L., 2014. Physical therapists' level of McKenzie education, functional outcomes, and utilization in patients with low back pain. *journal of orthopaedic & sports physical therapy*, *44*(12), pp.925-936.

Drummond, M.F., Sculpher, M.J., Claxton, K., Stoddart, G.L. and Torrance, G.W., 2015. *Methods for the economic evaluation of health care programmes*. Oxford university press..

Hart, D.L., Connolly, J.B., 2006. Pay-for-performance for physical therapy and occupational therapy: Medicare Part B Services. Final report. Grant 18-P-93066/9-01. Health & Human Services/Centers for https://www.cms.gov/Medicare/Billing/TherapyServices/downloads/P4PFinalReport06-01-06.pdf Medicare & Medicaid Services; 2006. Available at: Accessed 27/10/21

Hill, J.C., Whitehurst, D.G., Lewis, M., Bryan, S., Dunn, K.M., Foster, N.E., Konstantinou, K., Main, C.J., Mason, E., Somerville, S. and Sowden, G., 2011. Comparison of stratified primary care management for low back pain with current best practice (STarT Back): a randomised controlled trial. *The Lancet*, 378(9802), pp.1560-1571.

Kirkham, J.J., Davis, K., Altman, D.G., Blazeby, J.M., Clarke, M., Tunis, S. and Williamson, P.R., 2017. Core outcome Set-STAndards for development: the COS-STAD recommendations. *PLoS medicine*, *14*(11), p.e1002447.

Kooijman, M.K., Barten, D.J.A., Swinkels, I.C., Kuijpers, T., de Bakker, D., Koes, B.W. and Veenhof, C., 2015. Pain intensity, neck pain and longer duration of complaints predict poorer outcome in patients with shoulder pain–a systematic review. *BMC musculoskeletal disorders*, 16(1), p.288.

Kyte, D.G., Calvert, M., Van der Wees, P.J., Ten Hove, R., Tolan, S. and Hill, J.C., 2015. An introduction to patient-reported outcome measures (PROMs) in physiotherapy. *Physiotherapy*, *101*(2), pp.119-125.

Mallen, C.D., Peat, G., Thomas, E., Dunn, K.M. and Croft, P.R., 2007. Prognostic factors for musculoskeletal pain in primary care: a systematic review. *British Journal of General Practice*, *57*(541), pp.655-661.

Mallen, C.D., Thomas, E., Belcher, J., Rathod, T., Croft, P. and Peat, G., 2013. Point-of-care prognosis for common musculoskeletal pain in older adults. *JAMA internal medicine,* 173(12), pp.1119-1125.

Mogyarosy, Z., Smith, P., 2005. The main methodological issues in costing health care services A literature review. CHE Research Paper 7. P 43. Available at: https://www.york.ac.uk/media/che/documents/papers/researchpapers/rp7_Methodological_issues_in_costing_health_care_services.pdf Accessed 27/10/21

Moons, K.G., Altman, D.G., Vergouwe, Y. and Royston, P., 2009. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ*, *338*.

Nelson, E.C., Dixon-Woods, M., Batalden, P.B., Homa, K., Van Citters, A.D., Morgan, T.S., Eftimovska, E., Fisher, E.S., Ovretveit, J., Harrison, W. and Lind, C., 2016. Patient focused registries can improve health, care, and science. *BMJ*, 354, p.i3319.

NHS Digital 2020. Provisional Patient Reported Outcome Measures (PROMs) in England for Hip and Knee Replacement Procedures (April 2019 to March 2020) Available at: https://digital.nhs.uk/data-and-information/publications/statistical/patient-reported-outcome-measures-proms/hip-and-knee-replacement-procedures-april-2019-to-march-2020 Accessed 27/10/21

NHS England, 2014. NHS Five Year Forward View. Available at: https://www.england.nhs.uk/publication/nhs-five-year-forward-view/ Accessed 27/10/21

NHS England 2019. NHS RightCare MSK Focus Packs. Available at: http://tools.england.nhs.uk/cfv2016/msk/atlas.html Accessed 27/10/21

NHS Mandate, 2017. The Government's mandate to NHS England for 2017-18: A mandate from the Government to NHS England: April 2017 to March 2018. Available at: https://www.gov.uk/government/publications/nhs-mandate-2017-to-2018 Accessed 27/10/21

Phillips, S.D., Kramer, T.L., Compton, S.N., Burns, B.J. and Robbins, J.M., 2003. Case-mix adjustment of adolescent mental health treatment outcomes. *The journal of behavioural health services & research*, *30*(1), pp.125-136.

Porter, M.E. and Teisberg, E.O., 2006. *Redefining health care: creating value-based competition on results*. Harvard business press.

Porter, M.E., 2009. A strategy for health care reform—toward a value-based system. *New England Journal of Medicine,* 361(2), pp.109-112.

Rodeghero, J., Wang, Y.C., Flynn, T., Cleland, J.A., Wainner, R.S. and Whitman, J.M., 2015. The impact of physical therapy residency or fellowship education on clinical outcomes for patients with musculoskeletal conditions. *journal of orthopaedic & sports physical therapy*, *45*(2), pp.86-96.

Rolfson, O., Wissig, S., van Maasakkers, L., Stowell, C., Ackerman, I., Ayers, D., Barber, T., Benzakour, T., Bozic, K., Budhiparama, N. and Caillouette, J., 2016. Defining an international standard set of outcome measures for patients with hip or knee osteoarthritis: consensus of the international consortium for health outcomes measurement hip and knee osteoarthritis working group. *Arthritis care & research*, *68*(11), pp.1631-1639.

Royal College of Physicians, 2017a. National Hip Fracture Database (NHFD), National Report (2017). Available at: https://www.nhfd.co.uk/2017report Accessed 27/10/21

Royal College of Physicians, 2017b. Sentinel Stroke National Audit Programme (SSNAP). Available at: https://www.strokeaudit.org/Documents/National/Clinical/AprJul2017/AprJul2017-CCGLHBPublicReport.aspx Accessed 27/10/21

Siemens, D.R. and Booth, C.M., 2017. Benchmarking our urological care: It's just the beginning. *Canadian Urological Association Journal,* 11(8), p.223.

Smith, A.J., Dieppe, P., Vernon, K., Porter, M. and Blom, A.W., 2012. Failure rates of stemmed metal-on-metal hip replacements: analysis of data from the National Joint Registry of England and Wales. *The Lancet*, 379(9822), pp.1199-1204.

Struyf, F., Geeraerts, J., Noten, S., Meeus, M. and Nijs, J., 2016. A multivariable prediction model for the chronification of non-traumatic shoulder pain: A systematic review. *Pain physician*, 19(2), pp.1-10.

Van der Wees, P.J., Nijhuis-van Der Sanden, M.W., Ayanian, J.Z., Black, N., Westert, G.P. and Schneider, E.C., 2014. Integrating the Use of Patient-Reported Outcomes for Both Clinical Practice and Performance Measurement: Views of Experts from 3 Countries. *The Milbank Quarterly*, 92(4), pp.754-775.

Verkerk, K., Luijsterburg, P.A., Miedema, H.S., Pool-Goudzwaard, A. and Koes, B.W., 2012. Prognostic factors for recovery in chronic nonspecific low back pain: a systematic review. *Physical therapy*, 92(9), pp.1093-1108.

Versus Arthritis 2019. State of musculoskeletal health 2019; Arthritis and other musculoskeletal conditions in numbers. Available at: https://www.versusarthritis.org/media/14594/state-of-musculoskeletal-health-2019.pdf Accessed 27/10/21

Vos, T., Abajobir, A.A., Abate, K.H., Abbafati, C., Abbas, K.M., Abd-Allah, F., Abdulkader, R.S., Abdulle, A.M., Abebo, T.A., Abera, S.F. and Aboyans, V., 2017. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet*, *390*(10100), pp.1211-1259.

Werneke, M.W., Edmond, S., Deutscher, D., Ward, J., Grigsby, D., Young, M., McGill, T., McClenahan, B., Weinberg, J. and Davidow, A.L., 2016. Effect of adding McKenzie syndrome, centralization, directional preference, and psychosocial classification variables to a risk-adjusted model predicting functional status outcomes for patients with lumbar impairments. *Journal of Orthopaedic & Sports Physical Therapy*, 46(9), pp.726-741.

Whittle, R., Royle, K.L., Jordan, K.P., Riley, R.D., Mallen, C.D. and Peat, G., 2017. Prognosis research ideally should measure time-varying predictors at their intended moment of use. *Diagnostic and Prognostic Research*, 1(1), p.1.

Williamson, P.R., Altman, D.G., Bagley, H., Barnes, K.L., Blazeby, J.M., Brookes, S.T., Clarke, M., Gargon, E., Gorst, S., Harman, N. and Kirkham, J.J., 2017. The COMET handbook: version 1.0. *Trials*, *18*(3), pp.1-50.

Wind, A. and van Harten, W.H., 2017. Benchmarking specialty hospitals, a scoping review on theory and practice. *BMC health services research*, 17(1), p.245.

Chapter 2

Black, N., 2013. Patient reported outcome measures could help transform healthcare. *British Medical Journal (Online).* 346.

Braeken, A.M., Lochhaas-Gerlach, J.A., Gollish, J.D., MYĹES, J.D. and Mackenzie, T.A., 1997. Determinants of 6—12 Month Postoperative Functional Status and Pain After Elective Total Hip Replacement. *International Journal for Quality in Health Care*, *9*(6), pp.413-418.

Browne, J., Jamieson, L., Lewsey, J., van der Meulen, J., Black, N., Cairns, J., Lamping, D., Smith, S., Copley, L., Horrocks, J., 2007. Patient Reported Outcome Measures (PROMs) in elective surgery-report to the department of health. Health Services Research Unit, London School of Hygiene & Tropical Medicine & Clinical Effectiveness Unit, Royal College of Surgeons of England.

Burstin, H.R., Lipsitz, S.R. and Brennan, T.A., 1992. Socioeconomic status and risk for substandard medical care. *Jama*, *268*(17), pp.2383-2387.

Childs, J.D., Harman, J.S., Rodeghero, J.R., Horn, M. and George, S.Z., 2014. Implications of practice setting on clinical outcomes and efficiency of care in the delivery of physical therapy services. *journal of orthopaedic & sports physical therapy*, *44*(12), pp.955-963.

Critical Appraisal Skills Programme, 2017. CASP (Cohort Study) Checklist. Available at: http://docs.wixstatic.com/ugd/dded87_5ad0ece77a3f4fc9bcd3665a7d1fa91f.pdf Accessed: 01/11/21.

Cohen, J., 1988. *Statistical power analysis for the behavioral sciences*. 2nd ed. pp 79-82.

Coles, J. PROMs risk adjustment methodology guide for general surgery and orthopaedic procedures. Available at: https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2013/07/proms-ris-adj-meth-sur-orth.pdf Accessed 27/10/21

Department of Health., 2012. PROMs in England; the case-mix adjustment methodology. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/216507/dh_133449.pdf Accessed 27/10/21

Department of Health Website, 2017. [cited May 2017] Available at: https://www.gov.uk/government/organisations/department-of-health-and-social-care Accessed 01/11/21

Deutscher, D., Werneke, M.W., Gottlieb, D., Fritz, J.M. and Resnik, L., 2014. Physical therapists' level of McKenzie education, functional outcomes, and utilization in patients with low back pain. *journal of orthopaedic & sports physical therapy*, *44*(12), pp.925-936.

Devlin, N., Appleby, J., 2010. Getting the Most out of PROMs: Putting Health Outcomes at the Heart of NHS Decision Making, London: The King's Fund. Available at: https://www.kingsfund.org.uk/sites/default/files/Getting-the-most-out-of-PROMs-Nancy-Devlin-John-Appleby-Kings-Fund-March-2010.pdf Accessed 27/10/21.

Fanuele, J.C., Birkmeyer, N.J., Abdu, W.A., Tosteson, T.D. and Weinstein, J.N., 2000. The impact of spinal problems on the health status of patients: have we underestimated the effect?. *Spine*, *25*(12), pp.1509-1514.

Focus on Therapeutic Outcomes Inc. (FOTO Inc.) 2018 [cited August 2018] Available at: https://www.fotoinc.com/about-foto Accessed 01/11/21

Gomes, M., Gutacker, N., Bojke, C. and Street, A., 2016. Addressing missing data in patient-reported outcome measures (PROMS): Implications for the use of PROMS for comparing provider performance. *Health economics*, *25*(5), pp.515-528.

Gozalo, P.L., Resnik, L.J. and Silver, B., 2016. Benchmarking outpatient rehabilitation clinics using functional status outcomes. *Health services research*, *51*(2), pp.768-789.

Grigsby, J., Kaye, K., Kowalsky, J. and Kramer, A.M., 2002. Relationship between functional status and the capacity to regulate behavior among elderly persons following hip fracture. *Rehabilitation Psychology*, *47*(3), p.291.

Gutacker, N., Bojke, C., Daidone, S., Devlin, N. and Street, A., 2012. Analysing hospital variation in health outcome at the level of EQ-5D dimensions. CHE Research Paper, no. 74, Centre for Health Economics, University of York, York, UK.

Hart, D.L., Connolly, J.B., 2006. Pay-for-performance for physical therapy and occupational therapy: Medicare Part B Services. Final report. Grant. Jun 1:9-01. Available at: https://www.cms.gov/Medicare/Billing/TherapyServices/downloads/P4PFinalReport06-01-06.pdf Accessed 01/11/21

Hart, D.L., Werneke, M.W., Deutscher, D., George, S.Z. and Stratford, P.W., 2011a. Effect of fear-avoidance beliefs of physical activities on a model that predicts risk-adjusted functional status outcomes in patients treated for a lumbar spine dysfunction. *Journal of orthopaedic & sports physical therapy*, *41*(5), pp.336-345.

Hart, D.L., Werneke, M.W., Deutscher, D., George, S.Z., Stratford, P.W. and Mioduski, J.E., 2011b. Using intake and change in multiple psychosocial measures to predict functional status outcomes in people with lumbar spine syndromes: a preliminary analysis. *Physical therapy*, *91*(12), pp.1812-1825.

Higgins, J.P.T., and Green, S. (editors). 2011. *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0:130 [updated March 2011]. The Cochrane Collaboration, 2011. Available at: www.handbook.cochrane.org Accessed 01/11/21

Health and Social Care Information Centre (HSCIC), 2014. Finalised Patient Reported Outcome Measures (PROMs) in England: April 2012 to March 2013. Available at: https://www.gov.uk/government/statistics/patient-reported-outcome-measures-proms-in-england-finalised-april-2012-to-march-2013 Accessed 01/11/21

Iezzoni, L.I., 2009. 3. 1 Risk adjustment for performance measurement. *Performance measurement for health system improvement: experiences, challenges and prospects*, *251*.

Judge, A., Javaid, M.K., Arden, N.K., Cushnaghan, J., Reading, I., Croft, P., Dieppe, P.A. and Cooper, C., 2012. Clinical tool to identify patients who are most likely to achieve long-term improvement in physical function after total hip arthroplasty. *Arthritis care & research*, *64*(6), pp.881-889.

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P. and Stewart, L.A., 2015. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic reviews*, *4*(1), pp.1-9.

NHS England, 2013. Patient reported outcome measures in England: Update to reporting and case-mix adjusting hip and knee procedure data. Available at: https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2013/10/proms-meth-prim-revis.pdf Accessed 27/10/21

NHS England. 2016. Bite-size guide to patient insight: The National Patient Reported Outcome Measures PROMS Programme. Available at: https://www.england.nhs.uk/wp-content/uploads/2018/08/proms-guide-aug-18-v3.pdf Accessed 27/10/21

NICE Website, 2017. [cited May 2017] Available at: www.evidence.nhs.uk Accessed 01/11/21

Nuttall, D., Parkin, D. and Devlin, N., 2015. Inter-provider comparison of patient-reported outcomes: developing an adjustment to account for differences in patient case mix. *Health economics*, *24*(1), pp.41-54.

Petrie, A., and Sabin, C., 2013. *Medical statistics at a glance.* John Wiley & Sons; 2013 Nov 8. 84.

Resnik, L. and Hart, D.L., 2003. Using clinical outcomes to identify expert physical therapists. *Physical Therapy*, *83*(11), pp.990-1002.

Resnik, L., Liu, D., Mor, V. and Hart, D.L., 2008. Predictors of physical therapy clinic performance in the treatment of patients with low back pain syndromes. *Physical therapy*, *88*(9), pp.989-1004.

Resnik, L., Gozalo, P. and Hart, D.L., 2011. Weighted index explained more variance in physical function than an additively scored functional comorbidity scale. *Journal of clinical epidemiology*, *64*(3), pp.320-330.

Rodeghero, J.R., Cook, C.E., Cleland, J.A., Mintken, P.E., 2015. Risk stratification of patients with low back pain seen in physical therapy practice. *Manual therapy.* Dec 1;20(6):855-60.

Schäfer, T., Krummenauer, F., Mettelsiefen, J., Kirschner, S. and Günther, K.P., 2010. Social, educational, and occupational predictors of total hip replacement outcome. *Osteoarthritis and cartilage*, *18*(8), pp.1036-1042.

Vasseljen, O., Woodhouse, A., Bjørngaard, J.H. and Leivseth, L., 2013. Natural course of acute neck and low back pain in the general population: the HUNT study. *PAIN®*, *154*(8), pp.1237-1244.

Von Elm, E., Altman, D.G., Egger, M., Pocock, S.J., Gøtzsche, P.C. and Vandenbroucke, J.P., 2007. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Bulletin of the World Health Organization*, *85*, pp.867-872.

Ware, J.E., 1993. Scoring the SF-36. *SF-36. Health Survey: Manual and Interpretation Guide*.

Werneke, M. and Hart, D.L., 2001. Centralization phenomenon as a prognostic factor for chronic low back pain and disability. *Spine*, *26*(7), pp.758-764.

Werneke, M.W., Hart, D.L., Resnik, L., Stratford, P.W. and Reyes, A., 2008. Centralization: prevalence and effect on treatment outcomes using a standardized operational definition and measurement method. *journal of orthopaedic & sports physical therapy*, *38*(3), pp.116-125.

Werneke, M.W., Hart, D.L., George, S.Z., Stratford, P.W., Matheson, J.W. and Reyes, A., 2009. Clinical outcomes for patients classified by fear-avoidance beliefs and centralization phenomenon. *Archives of physical medicine and rehabilitation*, *90*(5), pp.768-777.

Werneke, M.W., Hart, D.L., George, S.Z., Deutscher, D. and Stratford, P.W., 2011a. Change in psychosocial distress associated with pain and functional status outcomes in patients with lumbar impairments referred to physical therapy services. *journal of orthopaedic & sports physical therapy*, *41*(12), pp.969-980.

Werneke, M.W., Hart, D.L., Cutrone, G., Oliver, D., McGill, M.T., Weinberg, J., Grigsby, D., Oswald, W. and Ward, J., 2011b. Association between directional preference and centralization in patients with low back pain. *journal of orthopaedic & sports physical therapy*, *41*(1), pp.22-31.

Werneke, M.W., Edmond, S., Deutscher, D., Ward, J., Grigsby, D., Young, M., McGill, T., McClenahan, B., Weinberg, J. and Davidow, A.L., 2016. Effect of adding McKenzie syndrome, centralization, directional preference, and psychosocial classification variables to a risk-adjusted model predicting functional status outcomes for patients with lumbar impairments. *journal of orthopaedic & sports physical therapy*, *46*(9), pp.726-741.

Whittle, R., Royle, K.L., Jordan, K.P., Riley, R.D., Mallen, C.D. and Peat, G., 2017. Prognosis research ideally should measure time-varying predictors at their intended moment of use. *Diagnostic and prognostic research*, *1*(1), pp.1-9.

Yen, S.C., Corkery, M.B., Chui, K.K., Manjourides, J., Wang, Y.C. and Resnik, L.J., 2015. Risk adjustment for lumbar dysfunction: comparison of linear mixed models with and without inclusion of between-clinic variation as a random effect. *Physical therapy*, *95*(12), pp.1692-1702.

Chapter 3

Altman, D. G., & Riley, R. D., 2005. Primer: an evidence-based approach to prognostic markers. *Nature Reviews Clinical Oncology*, *2*(9), 466.

Aromataris, E., Fernandez, R. S., Godfrey, C., Holly, C., Khalil, H., & Tungpunkom, P., 2014. Methodology for JBI umbrella reviews. In Joanna Briggs Institute reviewer's manual (2014). The Joanna Briggs Institute, Adelaide.

Aromataris, E., Fernandez, R., Godfrey, C. M., Holly, C., Khalil, H., & Tungpunkom, P., 2015. Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. *International journal of evidence-based healthcare*, *13*(3), 132-140.

Artus, M., Campbell, P., Mallen, C. D., Dunn, K. M., & van der Windt, D. A., 2017. Generic prognostic factors for musculoskeletal pain in primary care: a systematic review. *BMJ open*, *7*(1), e012901.

Bastick, A. N., Runhaar, J., Belo, J. N., & Bierma-Zeinstra, S. M., 2015. Prognostic factors for progression of clinical osteoarthritis of the knee: a systematic review of observational studies. *Arthritis research & therapy*, *17*(1), 152.

Becker, L., & Oxman, A., 2011. Chapter 22: Overviews of reviews In: Higgins JPT, Green S (editors), Cochrane Handbook for Systematic Reviews of Interventions Version 510 (updated March 2011) The Cochrane Collaboration, 2011. *Available: wwwcochrane-handbookorg*.

Browne, J., Jamieson, L., Lewsey, J., van der Meulen, J., Black, N., Cairns, J., Lamping, D., Smith, S., Copley, L., Horrocks, J., 2007. Patient Reported Outcome Measures (PROMs) in elective surgery-report to the department of health. Health Services Research Unit, London School of Hygiene & Tropical Medicine & Clinical Effectiveness Unit, Royal College of Surgeons of England.

Bruls, V. E., Bastiaenen, C. H., & de Bie, R. A., 2015. Prognostic factors of complaints of arm, neck, and/or shoulder: a systematic review of prospective cohort studies. *Pain*, *156*(5), 765-788.

Buirs, L. D., Van Beers, L. W. A. H., Scholtes, V. A. B., Pastoors, T., Sprague, S., & Poolman, R. W., 2016. Predictors of physical functioning after total hip arthroplasty: a systematic review. *BMJ open*, *6*(9), e010725.

Burgess, R., Bishop, A., Lewis, M., & Hill, J., 2018. Models used for case-mix adjustment of patient reported outcome measures (PROMs) in musculoskeletal healthcare: A systematic review of the literature. *Physiotherapy*. https://doi.org/10.1016/j.physio.2018.10.002

Burstin, H.R., Lipsitz, S.R. and Brennan, T.A., 1992. Socioeconomic status and risk for substandard medical care. *Jama*, *268*(17), pp.2383-2387.

Campbell, P., Hill, J.C., Protheroe, J., Afolabi, E.K., Lewis, M., Beardmore, R., Hay, E.M., Mallen, C.D., Bartlam, B., Saunders, B. and van der Windt, D.A., 2016. Keele Aches and Pains Study protocol: validity, acceptability, and feasibility of the Keele STarT MSK tool for subgrouping musculoskeletal patients in primary care. *Journal of pain research*, *9*, p.807.

Chester, R., Shepstone, L., Daniell, H., Sweeting, D., Lewis, J., & Jerosch-Herold, C., 2013. Predicting response to physiotherapy treatment for musculoskeletal shoulder pain: a systematic review. *BMC musculoskeletal disorders*, *14*(1), 203.

Coles, J., 2010. PROMs risk adjustment methodology guide for general surgery and orthopaedic procedures. Northgate. Available at: https://www.england.nhs.uk/statistics/wpcontent/uploads/sites/2/2013/07/proms-ris-adj-meth-sur-orth.pdf Accessed 01/11/21

Comer, C., Glover, J., Richardson, J., Yaseen, R., Foster, S., Wolfenden, N.M. and Hughes, G.J., 2016. Stratification of Treatment in a Community-Based Musculoskeletal Service: A Mixed-Methods Study to Assess Predictors of Requiring Complex Care. *Archives of physical medicine and rehabilitation*, *97*(6), pp.900-911.

Deeks, J.J., Dinnes, J., D'Amico, R., Sowden, A.J., Sakarovitch, C., Song, F., Petticrew, M. and Altman, D.G., 2003. Evaluating non-randomised intervention studies. *Health technology assessment (Winchester, England)*, *7*(27), pp.iii-173.

Department of Health., 2012. PROMs in England; the case-mix adjustment methodology. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/216507/dh_133449.pdf Accessed 27/10/21

de Rooij, M., van Der Leeden, M., Heymans, M.W., Holla, J.F., Häkkinen, A., Lems, W.F., Roorda, L.D., Veenhof, C., Sanchez-Ramirez, D.C., de Vet, H.C. and Dekker, J., 2016. Course and predictors of pain and physical functioning in patients with hip osteoarthritis: Systematic review and meta-analysis. *Journal of rehabilitation medicine*, *48*(3), 245-252.

Deutscher, D., Werneke, M.W., Hayes, D., Mioduski, J.E., Cook, K.F., Fritz, J.M., Woodhouse, L.J. and Stratford, P.W., 2018. Impact of risk adjustment on provider ranking for patients with low back pain receiving physical therapy. *journal of orthopaedic & sports physical therapy*, *48*(8), pp.637-648.

Devlin, N., Appleby, J., 2010. Getting the Most out of PROMs: Putting Health Outcomes at the Heart of NHS Decision Making, London: The King's Fund. Available at: https://www.kingsfund.org.uk/sites/default/files/Getting-the-most-out-of-PROMs-Nancy-Devlin-John-Appleby-Kings-Fund-March-2010.pdf Accessed 27/10/21.

Gozalo, P.L., Resnik, L.J. and Silver, B., 2016. Benchmarking outpatient rehabilitation clinics using functional status outcomes. *Health services research*, *51*(2), pp.768-789.

Gutacker, N., Bojke, C., Daidone, S., Devlin, N. and Street, A., 2012. Analysing hospital variation in health outcome at the level of EQ-5D dimensions. CHE Research Paper, no. 74, Centre for Health Economics, University of York, York, UK.

Guyatt, G., Oxman, A.D., Akl, E.A., Kunz, R., Vist, G., Brozek, J., Norris, S., Falck-Ytter, Y., Glasziou, P., DeBeer, H. and Jaeschke, R., 2011. GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. *Journal of clinical epidemiology*, *64*(4), pp.383-394.

Hancock, M., Herbert, R. D., & Maher, C. G., 2009. A guide to interpretation of studies investigating subgroups of responders to physical therapy interventions. *Physical therapy*, *89*(7), 698-704.

Harmelink, K. E., Zeegers, A. V., Hullegie, W., Hoogeboom, T. J., Nijhuis-van der Sanden, M. W., & Staal, J. B., 2017. Are there prognostic factors for one-year outcome after total knee arthroplasty? A systematic review. *The Journal of arthroplasty*, *32*(12), 3840-3853.

Hart, D.L., Connolly, J.B., 2006. Pay-for-performance for physical therapy and occupational therapy: Medicare Part B Services. Final report. Grant. Jun 1:9-01. Available at: https://www.cms.gov/Medicare/Billing/TherapyServices/downloads/P4PFinalReport06-01-06.pdf Accessed 01/11/21

Hart, D.L., Werneke, M.W., Deutscher, D., George, S.Z. and Stratford, P.W., 2011a. Effect of fear-avoidance beliefs of physical activities on a model that predicts risk-adjusted functional status outcomes in patients treated for a lumbar spine dysfunction. *Journal of orthopaedic & sports physical therapy*, *41*(5), pp.336-345.

Hart, D.L., Werneke, M.W., Deutscher, D., George, S.Z., Stratford, P.W. and Mioduski, J.E., 2011b. Using intake and change in multiple psychosocial measures to predict functional status outcomes in people with lumbar spine syndromes: a preliminary analysis. *Physical therapy*, *91*(12), pp.1812-1825.

Heerspink, F. O. L., Dorrestijn, O., van Raay, J. J., & Diercks, R. L., 2014. Specific patient-related prognostic factors for rotator cuff repair: a systematic review. *Journal of Shoulder and Elbow Surgery*, *23*(7), 1073-1080.

Hill, J. C., & Fritz, J. M., 2011. Psychosocial influences on low back pain, disability, and response to treatment. *Physical therapy*, *91*(5), 712-721.

Hill, J.C., Kang, S., Benedetto, E., Myers, H., Blackburn, S., Smith, S., Dunn, K.M., Hay, E., Rees, J., Beard, D. and Glyn-Jones, S., 2016. Development and initial cohort validation of the Arthritis Research UK Musculoskeletal Health Questionnaire (MSK-HQ) for use across musculoskeletal care pathways. *BMJ open*, 6(8), p.e012331.

Hofstede, S. N., Gademan, M. G., Vlieland, T. P. V., Nelissen, R. G., & Marang-van de Mheen, P. J., 2016. Preoperative predictors for outcomes after total hip replacement in patients with osteoarthritis: a systematic review. *BMC musculoskeletal disorders*, *17*(1), 212.

Iezzoni, L. I., 2009. 3. 1 Risk adjustment for performance measurement. *Performance measurement for health system improvement: experiences, challenges and prospects*, *251*.

Khangura, S., Konnyu, K., Cushman, R., Grimshaw, J., & Moher, D., 2012. Evidence summaries: the evolution of a rapid review approach. *Systematic reviews*, *1*(1), 10.

Kooijman, M. K., Barten, D. J. A., Swinkels, I. C., Kuijpers, T., de Bakker, D., Koes, B. W., & Veenhof, C., 2015. Pain intensity, neck pain and longer duration of complaints predict poorer outcome in patients with shoulder pain–a systematic review. *BMC musculoskeletal disorders*, *16*(1), 288.

Lungu, E., Maftoon, S., Vendittoli, P. A., & Desmeules, F., 2016a. A systematic review of preoperative determinants of patient-reported pain and physical function up to 2 years following primary unilateral total hip arthroplasty. *Orthopaedics & Traumatology: Surgery & Research*, *102*(3), 397-403.

Lungu, E., Vendittoli, P. A., & Desmeules, F., 2016b. Preoperative determinants of patient-reported pain and physical function levels following total knee arthroplasty: a systematic review. *The open orthopaedics journal*, *10*, 213.

Lunny, C., Brennan, S. E., McDonald, S., & McKenzie, J. E., 2017. Toward a comprehensive evidence map of overview of systematic review methods: paper 1—purpose, eligibility, search and data extraction. *Systematic reviews*, *6*(1), 231.

Magklara, E., Burton, C. R., & Morrison, V., 2014. Does self-efficacy influence recovery and well-being in osteoarthritis patients undergoing joint replacement? A systematic review. *Clinical rehabilitation*, *28*(9), 835-846.

Mallen, C. D., Peat, G., Thomas, E., Dunn, K. M., & Croft, P. R., 2007. Prognostic factors for musculoskeletal pain in primary care: a systematic review. *Br J Gen Pract*, *57*(541), 655-661.

McKillop, A. B., Carroll, L. J., & Battié, M. C., 2014. Depression as a prognostic factor of lumbar spinal stenosis: a systematic review. *The Spine Journal*, *14*(5), 837-846.

NHS England, 2013a. Patient reported outcome measures in England: Update to reporting and case-mix adjusting hip and knee procedure data. Available at: https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2013/10/proms-meth-prim-revis.pdf Accessed 27/10/21

NHS England, 2013b. Patent Reported Outcome Measures (PROMs) An alternative aggregation methodology for case-mix adjustment. Available at: http://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2013/07/proms-agg-meth-adju.pdf Accessed 27/10/21

Nuttall, D., Parkin, D. and Devlin, N., 2015. Inter-provider comparison of patient-reported outcomes: developing an adjustment to account for differences in patient case mix. *Health economics*, *24*(1), pp.41-54.

Office for National Statistics, 2018. Ethnicity pay gaps in Great Britain. Available at: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/articles/ethnicitypaygapsingreatbritain/2018 Accessed 27/10/21

Phillips, S. D., Kramer, T. L., Compton, S. N., Burns, B. J., & Robbins, J. M., 2003. Case-mix adjustment of adolescent mental health treatment outcomes. *The journal of behavioral health services & research*, *30*(1), 125-136.

Pinheiro, M.B., Ferreira, M.L., Refshauge, K., Maher, C.G., Ordoñana, J.R., Andrade, T.B., Tsathas, A. and Ferreira, P.H., 2016. Symptoms of depression as a prognostic factor for low back pain: a systematic review. *The Spine Journal*, *16*(1), pp.105-116.

Rebar, A. L., Stanton, R., Geard, D., Short, C., Duncan, M. J., & Vandelanotte, C., 2015. A meta-meta-analysis of the effect of physical activity on depression and anxiety in non-clinical adult populations. *Health psychology review*, *9*(3), 366-378.

Resnik, L., & Hart, D. L., 2003. Using clinical outcomes to identify expert physical therapists. *Physical therapy*, *83*(11), 990-1002.

Resnik, L., Gozalo, P. and Hart, D.L., 2011. Weighted index explained more variance in physical function than an additively scored functional comorbidity scale. *Journal of clinical epidemiology*, *64*(3), pp.320-330.

Riley, R.D., Hayden, J.A., Steyerberg, E.W., Moons, K.G., Abrams, K., Kyzas, P.A., Malats, N., Briggs, A., Schroter, S., Altman, D.G. and Hemingway, H., 2013. Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. *PLoS medicine*, *10*(2), p.e1001380.

Sackett, D.L., 2000. Therapy. *Evidence based medicine. How to practice and teach EBM*, 136-137.

Sephton, R., Hough, E., Roberts, S.A. and Oldham, J., 2010. Evaluation of a primary care musculoskeletal clinical assessment service: a preliminary study. *Physiotherapy*, *96*(4), pp.296-302.

Shea, B.J., Grimshaw, J.M., Wells, G.A., Boers, M., Andersson, N., Hamel, C., Porter, A.C., Tugwell, P., Moher, D. and Bouter, L.M., 2007. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC medical research methodology*, *7*(1), pp.1-7.

Struyf, F., Geeraerts, J., Noten, S., Meeus, M., & Nijs, J., 2016. A multivariable prediction model for the chronification of non-traumatic shoulder pain: a systematic review. *Pain Physician*, *19*(2), 1-10.

Verkerk, K., Luijsterburg, P.A., Miedema, H.S., Pool-Goudzwaard, A. and Koes, B.W., 2012. Prognostic factors for recovery in chronic nonspecific low back pain: a systematic review. *Physical therapy*, *92*(9), 1093-1108.

Von Korff, M., & Miglioretti, D. L., 2005. A prognostic approach to defining chronic pain. *Pain*, *117*(3), 304-313.

Werneke, M.W., Edmond, S., Deutscher, D., Ward, J., Grigsby, D., Young, M., McGill, T., McClenahan, B., Weinberg, J. and Davidow, A.L., 2016. Effect of adding McKenzie syndrome, centralization, directional preference, and psychosocial classification variables to a risk-adjusted model predicting functional status outcomes for patients with lumbar impairments. *journal of orthopaedic & sports physical therapy*, *46*(9), pp.726-741.

Wertli, M. M., Eugster, R., Held, U., Steurer, J., Kofmehl, R., & Weiser, S., 2014a. Catastrophizing—a prognostic factor for outcome in patients with low back pain: a systematic review. *The Spine Journal*, *14*(11), 2639-2657.

Wertli, M. M., Burgstaller, J. M., Weiser, S., Steurer, J., Kofmehl, R., & Held, U., 2014b. Influence of catastrophizing on treatment outcome in patients with nonspecific low back pain: a systematic review. *Spine*, *39*(3), 263-273.

Wertli, M. M., Rasmussen-Barr, E., Weiser, S., Bachmann, L. M., & Brunner, F., 2014c. The role of fear avoidance beliefs as a prognostic factor for outcome in patients with nonspecific low back pain: a systematic review. *The spine journal*, *14*(5), 816-836.

Wertli, M. M., Rasmussen-Barr, E., Held, U., Weiser, S., Bachmann, L. M., & Brunner, F., 2014d. Fear-avoidance beliefs—a moderator of treatment efficacy in patients with low back pain: a systematic review. *The Spine Journal*, *14*(11), 2658-2678.

Wilson, C.A., Roffey, D.M., Chow, D., Alkherayf, F. and Wai, E.K., 2016. A systematic review of preoperative predictors for postoperative clinical outcomes following lumbar discectomy. *The Spine Journal*, *16*(11), 1413-1422.

Woollard, J. D., Bost, J. E., Piva, S. R., Fitzgerald, G. K., Rodosky, M. W., & Irrgang, J. J., 2016. The relationship of preoperative factors to patient-reported outcome in rotator cuff repair: a systematic review. *Physical Therapy Reviews*, *21*(3-6), 138-150.

Yen, S.C., Corkery, M.B., Chui, K.K., Manjourides, J., Wang, Y.C. and Resnik, L.J., 2015. Risk adjustment for lumbar dysfunction: comparison of linear mixed models with and without inclusion of between-clinic variation as a random effect. *Physical therapy*, *95*(12), pp.1692-1702.

Chapter 4

ARUK. State of Musculoskeletal Health 2018: Arthritis and other musculoskeletal conditions in numbers. Available at: https://www.versusarthritis.org/about-arthritis/data-and-statistics/state-of-musculoskeletal-health-2019/ Accessed 01/11/21.

Barton, G.R., Sach, T.H., Jenkinson, C., Doherty, M., Avery, A.J. and Muir, K.R., 2009. Lifestyle interventions for knee pain in overweight and obese adults aged≥ 45: economic evaluation of randomised controlled trial. *Bmj*, *339*.

British Medical Association (BMA) 2021: *Models for paying providers of NHS services.* Available at: https://www.bma.org.uk/advice-and-support/nhs-delivery-and-workforce/funding/models-for-paying-providers-of-nhs-services Accessed 01/11/21

Bosmans, J.E., Pool, J.J., de Vet, H.C., van Tulder, M.W. and Ostelo, R.W., 2011. Is behavioral graded activity cost-effective in comparison with manual therapy for patients with subacute neck pain?: an economic evaluation alongside a randomized clinical trial. *Spine*, *36*(18), pp.E1179-E1186.

Bültmann, U., Sherson, D., Olsen, J., Hansen, C.L., Lund, T. and Kilsgaard, J., 2009. Coordinated and tailored work rehabilitation: a randomized controlled trial with economic evaluation undertaken with workers on sick leave due to musculoskeletal disorders. *Journal of occupational rehabilitation*, *19*(1), pp.81-93.

Carter, 2016. Operational productivity and performance in English NHS acute hospitals: Unwarranted variations. An independent report for the Department of Health by Lord Carter of Coles. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/499229/Operational_productivity_A.pdf Accessed 05/01/2021

Charles, J., Edwards, R.T., 2016. *A Guide to Health Economics for Those Working in Public Health*. Centre for Health Economics and Medicines Evaluation, Bangor University. Available at: https://cheme.bangor.ac.uk/documents/guide-handbook-en.pdf Accessed 01/11/21.

Chapman, C. Kern, A. 2010. *Costing in the National Health Service: from reporting to managing* (CIMA). July 2010. London

Chuang, L.H., Soares, M.O., Tilbrook, H., Cox, H., Hewitt, C.E., Aplin, J., Semlyen, A., Trewhela, A., Watt, I. and Torgerson, D.J., 2012. A pragmatic multicentered randomized controlled trial of yoga for chronic low back pain: economic evaluation. *Spine*, *37*(18), pp.1593-1601.

Curtis, L.A., & Burns, A., 2017. *Unit Costs of Health and Social Care 2017*. Personal Social Services Research Unit, University of Kent, Canterbury. https://doi.org/10.22024/UniKent/01.02/65559

Drummond, M.F., Sculpher, M.J., Claxton, K., Stoddart, G.L. and Torrance, G.W., 2015. *Methods for the economic evaluation of health care programmes*. Oxford university press.

Essex, H., Parrott, S., Atkin, K., Ballard, K., Bland, M., Eldred, J., Hewitt, C., Hopton, A., Keding, A., Lansdown, H. and Richmond, S., 2017. An economic evaluation of Alexander Technique lessons or

acupuncture sessions for patients with chronic neck pain: A randomized trial (ATLAS). *PloS one*, *12*(12), p.e0178918.

European Health Management Association (EHMA) 2007. HealthBASKET: Executive Summary and Policy Recommendations. Available at: https://cordis.europa.eu/docs/results/501/501588/126792701-6_en.pdf. Accessed 01/11/21.

Graves, N., Walker, D., Raine, R., Hutchings, A. and Roberts, J.A., 2002. Cost data for individual patients included in clinical studies: no amount of statistical analysis can compensate for inadequate costing methods. *Health economics*, *11*(8), pp.735-739.

Grieve, R., Cairns, J. and Thompson, S.G., 2010. Improving costing methods in multicentre economic evaluation: the use of multiple imputation for unit costs. *Health Economics*, *19*(8), pp.939-954.

Haines, T. and Bowles, K.A., 2017. Cost-effectiveness of using a motion-sensor biofeedback treatment approach for the management of sub-acute or chronic low back pain: economic evaluation alongside a randomised trial. *BMC musculoskeletal disorders*, *18*(1), pp.1-10.

Henchoz, Y., Pinget, C., Wasserfallen, J.B., Paillex, R., de Goumoëns, P., Norberg, M. and So, A.K.L., 2010. Cost-utility analysis of a three-month exercise programme vs usual care following multidisciplinary rehabilitation for chronic low back pain. *Journal of rehabilitation medicine*, *42*(9), pp.846-852.

Health Education England (HEE), 2020. A retrospective review of the influences, milestones, policies and practice developments in the First Contact MSK model: March 2020. Available at: http://arma.uk.net/wp-content/uploads/2020/04/MSK-march-2020-v2.pdf Accessed 22/11/21

HFMA, 2016. *Introductory guide - NHS finance:* 12th edition. HFMA. Bristol. pp 121-128 and 155-160.

Hobbs, F.R., Bankhead, C., Mukhtar, T., Stevens, S., Perera-Salazar, R., Holt, T. and Salisbury, C., 2016. Clinical workload in UK primary care: a retrospective analysis of 100 million consultations in England, 2007–14. *The Lancet*, *387*(10035), pp.2323-2330.

Hollinghurst, S., Sharp, D., Ballard, K., Barnett, J., Beattie, A., Evans, M., Lewith, G., Middleton, K., Oxford, F., Webley, F. and Little, P., 2008. Randomised controlled trial of Alexander technique lessons, exercise, and massage (ATEAM) for chronic and recurrent back pain: economic evaluation. *Bmj*, *337*.

Hollinghurst, S., Coast, J., Busby, J., Bishop, A., Foster, N.E., Franchini, A., Grove, S., Hall, J., Hopper, C., Kaur, S. and Montgomery, A.A., 2013. A pragmatic randomised controlled trial of 'PhysioDirect'telephone assessment and advice services for patients with musculoskeletal problems: economic evaluation. *BMJ open*, *3*(10), p.e003406.

Horngren, C.T., Datar, S.M., Foster, G., 2003. Cost Accounting: A Managerial Emphasis. 11th Ed. Prentice Hall. p. 29.

Jackson, D., 2012. Health economics made easy. Scion; Jan 31.

Jensen, C.E., Riis, A., Petersen, K.D., Jensen, M.B. and Pedersen, K.M., 2017. Economic evaluation of an implementation strategy for the management of low back pain in general practice. *Pain*, *158*(5), pp.891-899.

Johnston, K., Buxton, M.J., Jones, D.R. and Fitzpatrick, R., 2001. Collecting resource use data for costing in clinical trial. *The Advanced Handbook of Methods in Evidence Based Healthcare. London, England: Sage*, pp.215-230.

Jordan, K.P., Kadam, U.T., Hayward, R., Porcheret, M., Young, C. and Croft, P., 2010. Annual consultation prevalence of regional musculoskeletal problems in primary care: an observational study. *BMC musculoskeletal disorders*, *11*(1), pp.1-10.

Jowett, S., Crawshaw, D.P., Helliwell, P.S., Hensor, E.M., Hay, E.M. and Conaghan, P.G., 2013. Cost-effectiveness of exercise therapy after corticosteroid injection for moderate to severe shoulder pain due to subacromial impingement syndrome: a trial-based analysis. *Rheumatology*, *52*(8), pp.1485-1491.

Lamb, S.E., Hansen, Z., Lall, R., Castelnuovo, E., Withers, E.J., Nichols, V., Potter, R., Underwood, M.R. and Back Skills Training Trial investigators, 2010. Group cognitive behavioural treatment for low-back pain in primary care: a randomised controlled trial and cost-effectiveness analysis. *The Lancet*, *375*(9718), pp.916-923.

Lambeek, L.C., Bosmans, J.E., Van Royen, B.J., Van Tulder, M.W., Van Mechelen, W. and Anema, J.R., 2010. Effect of integrated care for sick listed patients with chronic low back pain: economic evaluation alongside a randomised controlled trial. *Bmj*, *341*.

Marra, C.A., Grubisic, M., Cibere, J., Grindrod, K.A., Woolcott, J.C., Gastonguay, L. and Esdaile, J.M., 2014. Cost-Utility Analysis of a Multidisciplinary Strategy to Manage Osteoarthritis of the Knee: Economic Evaluation of a Cluster Randomized Controlled Trial Study. *Arthritis care & research*, *66*(6), pp.810-816.

McKenna, C., Bojke, L., Manca, A., Adebajo, A., Dickson, J., Helliwell, P., Morton, V., Russell, I., Torgerson, D. and Watson, J., 2009. Shoulder acute pain in primary health care: is retraining GPs effective? The SAPPHIRE randomized trial: a cost–effectiveness analysis. *Rheumatology*, *48*(5), pp.558-563.

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P. and Stewart, L.A., 2015. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic reviews*, *4*(1), pp.1-9.

Monitor, 2015. Reference cost assurance programme: Findings from the 2014/15 audit. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/466112/Reference_cost_audit_report_final_v2.pdf Accessed 01/11/21

NHS England, 2014a. CCG Programme Budgeting Benchmarking Tool 2013/14. Available at: https://www.england.nhs.uk/prog-budgeting/ Accessed 01/11/21.

NHS England, 2014b. *NHS Five Year Forward View (FYFV).* Available at: https://www.england.nhs.uk/publication/nhs-five-year-forward-view/ Accessed 01/11/21

NHS England. 2016. Bite-size guide to patient insight: The National Patient Reported Outcome Measures PROMS Programme. Available at: https://www.england.nhs.uk/wp-content/uploads/2018/08/proms-guide-aug-18-v3.pdf Accessed 27/10/21

NHS England, 2017. *Transforming musculoskeletal and orthopaedic elective care services:* Available at: https://www.england.nhs.uk/wp-content/uploads/2017/11/msk-orthopaedic-elective-care-handbook-v2.pdf Accessed 01/11/21

NHS England, 2021a. The model health system. Available at: https://www.england.nhs.uk/applications/model-hospital/ Accessed 01/11/21

NHS England, 2021b. Approved Costing Guidance. Available at: https://www.england.nhs.uk/approved-costing-guidance/#ACG_2021 Accessed 01/11/21

NHS England and NHS Improvement, 2019. *Elective Care High Impact Interventions: First Contact Practitioner for MSK Services. May 2019.* Available at: https://www.england.nhs.uk/wp-content/uploads/2019/05/elective-care-high-impact-interventions-first-contact-practitioner-msk-services-specification.pdf Accessed 01/11/21

NHS England and NHS Improvement, 2020. Approved Costing Guidance – Standards Integrated information requirements. Available at: https://www.england.nhs.uk/wp-content/uploads/2020/08/INT_IRs_FINAL.pdf Accessed 01/11/21

NHS Improvement, 2016a. *Introduction to the Model Hospital Prototype Portal,* https://model.nhs.uk/downloads/ModelHospitalPortal_Guidance.pdf Accessed 24/06/19.

NHS Improvement, 2016b. *Costing transformation programme: Patient level costing; case for change. April 2016.* Available at: https://webarchive.nationalarchives.gov.uk/ukgwa/20200501112705mp_/https://improvement.nhs.uk/documents/36/CTP_PLICS_case_for_change.pdf Accessed 01/11/21

NHS Improvement, 2017. 2016/2017 reference costs and guidance. Available at: https://webarchive.nationalarchives.gov.uk/ukgwa/20200501111106/https://improvement.nhs.uk/resources/reference-costs/ Accessed 01/11/21

NHS Mandate, 2018. *The Government's mandate to NHS England for 2017-18: A mandate from the Government to NHS England:* April 2017 to March 2018. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/692140/NHSE_Mandate__2017-18_revised.pdf Accessed 01/11/21

NHS Rightcare, 2016. Focus packs for Cancer, Mental health and dementia, MSK and trauma – May 2016: https://www.england.nhs.uk/rightcare/products/ccg-data-packs/focus-packs/focus-packs-for-cancer-mental-health-and-dementia-msk-and-trauma-may-2016/ Accessed 01/11/21.

NHS RightCare, 2019. What is NHS RightCare? Available at: https://www.england.nhs.uk/rightcare/what-is-nhs-rightcare/ Accessed 25/11/2019

Pinto, D., Robertson, M.C., Abbott, J.H., Hansen, P., Campbell, A.J. and MOA Trial Team, 2013. Manual therapy, exercise therapy, or both, in addition to usual care, for osteoarthritis of the hip or knee. 2: economic evaluation alongside a randomized controlled trial. *Osteoarthritis and cartilage*, *21*(10), pp.1504-1513.

Porter, M.E., 2009. A strategy for health care reform—toward a value-based system. *N Engl J Med*, *361*(2), pp.109-112.

Porter, M.E., 2010. What is value in health care. *N Engl J Med*, *363*(26), pp.2477-2481.

Saha, S., Grahn, B., Gerdtham, U.G., Stigmar, K., Holmberg, S. and Jarl, J., 2019. Structured physiotherapy including a work place intervention for patients with neck and/or back pain in primary care: an economic evaluation. *The European Journal of Health Economics*, *20*(2), pp.317-327.

Smeets, R.J., Severens, J.L., Beelen, S., Vlaeyen, J.W. and Knottnerus, J.A., 2009. More is not always better: cost-effectiveness analysis of combined, single behavioral and single physical rehabilitation programs for chronic low back pain. *European Journal of Pain*, *13*(1), pp.71-81.

Struijs, J.N. and Baan, C.A., 2011. Integrating care through bundled payments—lessons from the Netherlands. *New England Journal of Medicine*, *364*(11), pp.990-991.

Urwin, M., Symmons, D., Allison, T., Brammah, T., Busby, H., Roxby, M., Simmons, A. and Williams, G., 1998. Estimating the burden of musculoskeletal disorders in the community: the comparative prevalence of symptoms at different anatomical sites, and the relation to social deprivation. *Annals of the rheumatic diseases*, *57*(11), pp.649-655.

van der Roer, N., van Tulder, M., van Mechelen, W. and de Vet, H., 2008. Economic evaluation of an intensive group training protocol compared with usual care physiotherapy in patients with chronic low back pain. *Spine*, *33*(4), pp.445-451.

Vermeulen, S.J., Heymans, M.W., Anema, J.R., Schellart, A.J., van Mechelen, W. and van der Beek, A.J., 2013. Economic evaluation of a participatory return-to-work intervention for temporary agency and unemployed workers sick-listed due to musculoskeletal disorders. *Scandinavian journal of work, environment & health*, pp.46-56.

Werner, E.L., Storheim, K., Løchting, I., Wisløff, T. and Grotle, M., 2016. Cognitive patient education for low back pain in primary care: a cluster randomized controlled trial and cost-effectiveness analysis. *Spine*, *41*(6), pp.455-462.

Whitehurst, D.G., Bryan, S., Lewis, M., Hay, E.M., Mullis, R. and Foster, N.E., 2015. Implementing stratified primary care management for low back pain: cost-utility analysis alongside a prospective, population-based, sequential comparison study. *Spine*, *40*(6), pp.405-414.

Chapter 5

Arthritis Research UK (ARUK), 2016. Recommended Musculoskeletal Indicator Set. Available at: https://www.versusarthritis.org/media/2125/recommended-msk-indicator-set-report.pdf Accessed 01/11/21

Berwick, D., 2013. A promise to learn – a commitment to act: improving the safety of patients in England. London: Department of Health. Department of Health and NHS Commissioning Board (2012) Compassion in practice – nursing, midwifery and care staff – our vision and strategy. London: Department of Health.

British Society of Rheumatology (BSR), 2019. National Early Inflammatory Arthritis Audit 1st Annual Report; 2019. Available at: https://www.rheumatology.org.uk/Portals/0/Documents/Practice_Quality/Audit/NEIA/2019/NEIA_Audit_report_October_2019.pdf?ver=2019-10-08-103326-710 Accessed 01/11/21

Burgess, R., Bishop, A., Lewis, M. and Hill, J., 2019. Models used for case-mix adjustment of patient reported outcome measures (PROMs) in musculoskeletal healthcare: A systematic review of the literature. *Physiotherapy*, *105*(2), pp.137-146.

Burgess, R., Mansell, G., Bishop, A., Lewis, M. and Hill, J., 2020a. Predictors of functional outcome in musculoskeletal healthcare: an umbrella review. *European Journal of Pain*, *24*(1), pp.51-70.

Burgess, R., Hall, J., Bishop, A., Lewis, M. and Hill, J., 2020b. Costing Methodology and Key Drivers of Health Care Costs Within Economic Analyses in Musculoskeletal Community and Primary Care Services: A Systematic Review of the Literature. *Journal of primary care & community health*, *11*, p.2150132719899763.

Campbell, P., Hill, J.C., Protheroe, J., Afolabi, E.K., Lewis, M., Beardmore, R., Hay, E.M., Mallen, C.D., Bartlam, B., Saunders, B. and van der Windt, D.A., 2016. Keele Aches and Pains Study protocol: validity, acceptability, and feasibility of the Keele STarT MSK tool for subgrouping musculoskeletal patients in primary care. *Journal of pain research,* 9, p.807.

Chartered Society of Physiotherapy (CSP), 2016. Outcome and experience measures. Available at: https://www.csp.org.uk/professional-clinical/research-evaluation/outcome-experience-measures Accessed 22/11/21

Chartered Society of Physiotherapy (CSP), 2020. National Evaluation of First Contact Practitioner (FCP) model of primary care. Available at: https://www.csp.org.uk/system/files/documents/2020-11/final_fcp_phase_3_national_evaluation_report.pdf Accessed 01/11/21

Clement, R.C., Welander, A., Stowell, C., Cha, T.D., Chen, J.L., Davies, M., Fairbank, J.C., Foley, K.T., Gehrchen, M., Hagg, O. and Jacobs, W.C., 2015. A proposed set of metrics for standardized outcome reporting in the management of low back pain. *Acta orthopaedica*, *86*(5), pp.523-533.

COMET Website, 2020. Available at: http://www.comet-initiative.org/Studies/Details/1551 Accessed 01/11/21

Coyle, J. and Williams, B., 2001. Valuing people as individuals: development of an instrument through a survey of person-centredness in secondary care. *Journal of Advanced Nursing*, *36*(3), pp.450-459.

Euroqol, 2021. EQ5D User Guide. Available at: https://euroqol.org/publications/user-guides/ Accessed 01/11/21

Diamond, I.R., Grant, R.C., Feldman, B.M., Pencharz, P.B., Ling, S.C., Moore, A.M. and Wales, P.W., 2014. Defining consensus: a systematic review recommends methodologic criteria for reporting of Delphi studies. *Journal of clinical epidemiology*, *67*(4), pp.401-409.

Elwyn, G., Barr, P.J., Grande, S.W., Thompson, R., Walsh, T. and Ozanne, E.M., 2013. Developing CollaboRATE: a fast and frugal patient-reported measure of shared decision making in clinical encounters. *Patient education and counseling*, *93*(1), pp.102-107.

Fennelly, O., Blake, C., Desmeules, F., Stokes, D. and Cunningham, C., 2018. Patient-reported outcome measures in advanced musculoskeletal physiotherapy practice: a systematic review. *Musculoskeletal Care*, *16*(1), pp.188-208.

Gibbons, E. and Fitzpatrick, R., 2018. An alternative approach to implementing patient-reported outcome measures. *Pilot and feasibility studies*, *4*(1), pp.1-9.

Hill, J.C., Whitehurst, D.G., Lewis, M., Bryan, S., Dunn, K.M., Foster, N.E., Konstantinou, K., Main, C.J., Mason, E., Somerville, S. and Sowden, G., 2011. Comparison of stratified primary care management for low back pain with current best practice (STarT Back): a randomised controlled trial. *The Lancet*, *378*(9802), pp.1560-1571.

Hill, J.C., Kang, S., Benedetto, E., Myers, H., Blackburn, S., Smith, S., Dunn, K.M., Hay, E., Rees, J., Beard, D. and Glyn-Jones, S., 2016. Development and initial cohort validation of the Arthritis Research UK Musculoskeletal Health Questionnaire (MSK-HQ) for use across musculoskeletal care pathways. *BMJ open*, *6*(8), p.e012331.

Hill, J., Garvin, S., Chen, Y., Cooper, V., Wathall, S., Bartlam, B., Saunders, B., Lewis, M., Protheroe, J., Chudyk, A. and Birkinshaw, H., 2020. Computer-based stratified primary care for musculoskeletal consultations compared with usual care: study protocol for the STarT MSK cluster randomized controlled trial. *JMIR research protocols*, *9*(7), p.e17939.

ICHOM, 2017. ICHOM Low Back Pain Data Collection Reference Guide. Available at: https://ichom.org/files/medical-conditions/low-back-pain/low-back-pain-reference-guide.pdf Accessed 27/10/21

ICHOM, 2017. ICHOM Hip and Knee Osteoarthritis Data Collection Reference Guide. Available at: https://ichom.org/files/medical-conditions/hip-knee-osteoarthritis/hip-knee-osteoarthritisreference-guide.pdf Accessed 27/10/21

Kirkham, J.J., Davis, K., Altman, D.G., Blazeby, J.M., Clarke, M., Tunis, S. and Williamson, P.R., 2017. Core outcome Set-STAndards for development: the COS-STAD recommendations. *PLoS medicine*, *14*(11), p.e1002447.

Kloppenburg, M., Bøyesen, P., Smeets, W., Haugen, I., Liu, R., Visser, W. and van der Heijde, D.M., 2014. Report from the OMERACT Hand Osteoarthritis Special Interest Group: advances and future research priorities. *The Journal of rheumatology*, *41*(4), pp.810-818.

Klokkerud, M., Dagfinrud, H., Uhlig, T., Dager, T.N., Furunes, K.A., Klokkeide, Å., Larsen, M., Nygård, S., Nylenna, S., Øie, L. and Kjeken, I., 2018. Developing and testing a consensus-based core set of outcome measures for rehabilitation in musculoskeletal diseases. *Scandinavian journal of rheumatology*, *47*(3), pp.225-234.

Lynch, A.D., Logerstedt, D.S., Grindem, H., Eitzen, I., Hicks, G.E., Axe, M.J., Engebretsen, L., Risberg, M.A. and Snyder-Mackler, L., 2015. Consensus criteria for defining 'successful outcome'after ACL injury and reconstruction: a Delaware-Oslo ACL cohort investigation. *British journal of sports medicine*, *49*(5), pp.335-342.

McRobert, C.J., Hill, J.C., Smale, T., Hay, E.M. and Van der Windt, D.A., 2018. A multi-modal recruitment strategy using social media and internet-mediated methods to recruit a multidisciplinary, international sample of clinicians to an online research study. *PLoS One*, *13*(7), p.e0200184.

National Joint Registry, 2020. 17th Annual Report 2020: Surgical data to 31 December 2019. Available at: https://reports.njrcentre.org.uk/Portals/0/PDFdownloads/NJR%2017th%20Annual%20Report%202020.pdf Accessed 01/11/21

NHS England, 2020. Finalised Patient Reported Outcome Measures (PROMs) in England for Hip & Knee Replacements, April 2018 – March 2019. Available at: https://digital.nhs.uk/data-and-information/publications/statistical/patient-reported-outcome-measures-proms/finalised-hip--knee-replacements-april-2018---march-2019 Accessed 01/11/21

NHS England Website 2019a. Friends and Family Test: Available at: https://www.england.nhs.uk/fft/ Accessed 01/11/21

NHS England, 2019b. Shared Decision Making: Summary Guide. https://www.england.nhs.uk/wp-content/uploads/2019/01/shared-decision-making-summary-guide-v1.pdf

NHS England and NHS Improvement, 2020. Primary and community musculoskeletal adult services: Restoration principles v2.0. Available at: https://www.boa.ac.uk/resources/primary-and-community-musculoskeletal-adult-services-restoration-principles.html

NHS Mandate (2019) The Government's revised mandate to NHS England for 2018-19: A mandate from the Government to NHS England: April 2018 to March 2019. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/803111/revised-mandate-to-nhs-england-2018-to-2019.pdf

NICE 2016. Low back pain and sciatica in over 16s: assessment and management. Available at: https://www.nice.org.uk/guidance/ng59 Accessed 23/11/21

Price, A.J., Ogollah, R., Kang, S., Hay, E., Barker, K.L., Benedetto, E., Smith, S., Smith, J., Galloway, J.B., Ellis, B. and Rees, J., 2019. Determining responsiveness and meaningful changes for the

Musculoskeletal Health Questionnaire (MSK-HQ) for use across musculoskeletal care pathways. *BMJ open*, *9*(10), p.e025357.

Rolfson, O., Wissig, S., van Maasakkers, L., Stowell, C., Ackerman, I., Ayers, D., Barber, T., Benzakour, T., Bozic, K., Budhiparama, N. and Caillouette, J., 2016. Defining an international standard set of outcome measures for patients with hip or knee osteoarthritis: consensus of the international consortium for health outcomes measurement hip and knee osteoarthritis working group. *Arthritis care & research*, *68*(11), pp.1631-1639.

Royal College of Physicians (RCP). National Hip Fracture Database (NHFD) Annual Report 2019. London: RCP, 2019. Available at: https://www.nhfd.co.uk/files/2019ReportFiles/NHFD_2019_Annual_Report_v101.pdf

Slater, H. and Briggs, A.M., 2017. Models of Care for musculoskeletal pain conditions: driving change to improve outcomes. Pain Management. 7(5), pp.351-357

Stratford, P., Gill, C., Westaway, M. and Binkley, J., 1995. Assessing disability and change on individual patients: a report of a patient specific measure. *Physiotherapy canada*, *47*(4), pp.258-263.

Vagias WM. Likert-type scale response anchors. clemson international institute for tourism. & Research Development, Department of Parks, Recreation and Tourism Management, Clemson University. 2006:4-5.

Williamson, P.R., Altman, D.G., Bagley, H., Barnes, K.L., Blazeby, J.M., Brookes, S.T., Clarke, M., Gargon, E., Gorst, S., Harman, N. and Kirkham, J.J., 2017. The COMET handbook: version 1.0. *Trials*, *18*(3), pp.1-50.

Chapter 6

Baker, P.N., Deehan, D.J., Lees, D., Jameson, S., Avery, P.J., Gregg, P.J. and Reed, M.R., 2012. The effect of surgical factors on early patient-reported outcome measures (PROMS) following total knee replacement. *The Journal of bone and joint surgery. British volume*, *94*(8), pp.1058-1066.

Bryman, A. and Cramer, D., 2002. *Quantitative data analysis with SPSS release 10 for Windows: A guide for social scientists*. Routledge.

Burgess, R., Bishop, A., Lewis, M. and Hill, J., 2019. Models used for case-mix adjustment of patient reported outcome measures (PROMs) in musculoskeletal healthcare: A systematic review of the literature. *Physiotherapy*, *105*(2), pp.137-146.

Burgess, R., Mansell, G., Bishop, A., Lewis, M. and Hill, J., 2020. Predictors of functional outcome in musculoskeletal healthcare: an umbrella review. *European Journal of Pain*, *24*(1), pp.51-70.

Coles, J., 2010. PROMs risk adjustment methodology guide for general surgery and orthopaedic procedures. Northgate. Available at: https://www.england.nhs.uk/statistics/wpcontent/uploads/sites/2/2013/07/proms-ris-adj-meth-sur-orth.pdf Accessed 22/11/21

Department of Health (DoH), 2012. PROMs in England; the case-mix adjustment methodology. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/216507/dh_133449.pdf Accessed 22/11/21

Deutscher, D., Werneke, M.W., Hayes, D., Mioduski, J.E., Cook, K.F., Fritz, J.M., Woodhouse, L.J. and Stratford, P.W., 2018. Impact of risk adjustment on provider ranking for patients with low back pain receiving physical therapy. *journal of orthopaedic & sports physical therapy*, *48*(8), pp.637-648.

Focus on Therapeutic Outcomes (FOTO) Website, 2021. https://www.fotoinc.com/

Hart, D.L., Connolly J.B., 2006. Pay-for-performance for physical therapy and occupational therapy: Medicare Part B Services. Final report. Grant 18-P-93066/9-01. Health & Human Services/Centers for Medicare & Medicaid Services. Available at: http://www.cms.hhs.gov/TherapyServices/downloads/P4PFinalReport06- 01-06.pdf Accessed 22/11/21

Hill, J.C., Kang, S., Benedetto, E., Myers, H., Blackburn, S., Smith, S., Dunn, K.M., Hay, E., Rees, J., Beard, D. and Glyn-Jones, S., 2016. Development and initial cohort validation of the Arthritis Research UK Musculoskeletal Health Questionnaire (MSK-HQ) for use across musculoskeletal care pathways. *BMJ open*, *6*(8), p.e012331.

Hill, J., Garvin, S., Chen, Y., Cooper, V., Wathall, S., Bartlam, B., Saunders, B., Lewis, M., Protheroe, J., Chudyk, A. and Birkinshaw, H., 2020. Computer-based stratified primary care for musculoskeletal consultations compared with usual care: study protocol for the STarT MSK cluster randomized controlled trial. *JMIR research protocols*, *9*(7), p.e17939.

Iezzoni, L.I., 2009. 3. 1 Risk adjustment for performance measurement. *Performance measurement for health system improvement: experiences, challenges and prospects*, *251*.

Krumholz, H.M., Brindis, R.G., Brush, J.E., Cohen, D.J., Epstein, A.J., Furie, K., Howard, G., Peterson, E.D., Rathore, S.S., Smith Jr, S.C. and Spertus, J.A., 2006. Standards for statistical models used for public reporting of health outcomes: an American Heart Association Scientific Statement from the Quality of Care and Outcomes Research Interdisciplinary Writing Group: cosponsored by the Council on Epidemiology and Prevention and the Stroke Council Endorsed by the American College of Cardiology Foundation. *Circulation*, *113*(3), pp.456-462.

Lang, T.A. and Altman, D.G., 2015. Basic statistical reporting for articles published in biomedical journals: the" Statistical Analyses and Methods in the Published Literature" or the SAMPL Guidelines. *Int J Nurs Stud*, *52*(1), pp.5-9.

Lutz, A.D., Brooks, J.M., Chapman, C.G., Shanley, E., Stout, C.E. and Thigpen, C.A., 2020. Risk Adjustment of the Modified Low Back Pain Disability Questionnaire and Neck Disability Index to Benchmark Physical Therapist Performance: Analysis From an Outcomes Registry. *Physical therapy*, *100*(4), pp.609-620.

NHS Digital, 2018. PROMs clinical case study: data informs clinical practice. Available at: PROMs clinical case study: data informs clinical practice - NHS Digital Accessed 22/11/21

NHS England, 2013. Patent Reported Outcome Measures (PROMs) An alternative aggregation methodology for case-mix adjustment. Available at: http://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2013/07/proms-agg-meth-adju.pdf Accessed 22/11/21

NHS England, 2016. Bite-size guide to patient insight: The National Patient Reported Outcome Measures PROMS Programme. Available at: https://www.england.nhs.uk/wp-content/uploads/2018/08/proms-guide-aug-18-v3.pdf Accessed 22/11/21

NHS England Website, 2021. Available at: https://www.england.nhs.uk/data-services/ Accessed 13/10/21

Nuttall, D., Parkin, D. and Devlin, N., 2015. Inter-provider comparison of patient-reported outcomes: developing an adjustment to account for differences in patient case mix. *Health economics*, *24*(1), pp.41-54.

Price, A.J., Ogollah, R., Kang, S., Hay, E., Barker, K.L., Benedetto, E., Smith, S., Smith, J., Galloway, J.B., Ellis, B. and Rees, J., 2019. Determining responsiveness and meaningful changes for the Musculoskeletal Health Questionnaire (MSK-HQ) for use across musculoskeletal care pathways. *BMJ open*, *9*(10), p.e025357.

Prince, 2012. M. Prince Epidemiology In; P. Wright, J. Stern, M. Phelan (Eds.), Core Psychiatry, Saunders Elsevier, Philadelphia (2012), pp. 126. https://doi.org/10.1016/B978-0-7020-3397-1.00009-4 Accessed 22/11/21

Scott, I.C., McCray, G., Lancaster, G., Foster, N.E. and Hill, J.C., 2020, October. Validation of the Musculoskeletal Health Questionnaire (MSK-HQ) in primary care patients with musculoskeletal pain. In *Seminars in Arthritis and Rheumatism* (Vol. 50, No. 5, pp. 813-820). WB Saunders.

Senaviratna, N.A.M.R. and Cooray, T.M.J.A., 2019. Diagnosing multicollinearity of logistic regression model. *Asian Journal of Probability and Statistics*, pp.1-9.

Siemens, D.R. and Booth, C.M., 2017. Benchmarking our urological care: It's just the beginning. *Canadian Urological Association Journal*, *11*(8), p.223.

van der Gaag W. H., Chiarotto, A., Heymans, M, W., et al. 2020. Developing clinical prediction models for nonrecovery in older patients seeking care for back pain: the back complaints in the elders prospective cohort study. *Pain*. DOI: 10.1097/j.pain.0000000000002161.

Chapter 7

Anhang Price R, Elliott MN, Zaslavsky AM, Hays RD, Lehrman WG, Rybowski L, et al. Examining the role of patient experience surveys in measuring health care quality. Medical care research and review: MCRR. 2014;71(5):522– 54. doi: 10.1177/ 1077558714541480.

ARMA Website 2021. Available at: http://arma.uk.net/bestmskhealth-programme-launch/ Accessed 10/09/21

Arthritis Research UK (ARUK). Recommended Musculoskeletal Indicator Set. 2016. Available at: https://www.versusarthritis.org/media/2125/recommended-msk-indicator-set-report.pdf Accessed 11/10/21

British Society of Rheumatology (2021). National Early Inflammatory Arthritis Audit – Second Annual Report. 8 May 2019 - 7 May 2020. Healthcare Quality Improvement Partnership (HQIP). Available at: https://www.rheumatology.org.uk/practice-quality/audits/neiaa Accessed 11/10/21

Briggs TWR (2012). Getting it right first time: improving the quality of orthopaedic care within the National Health Service in England. Available at: www.timbriggs-gettingitrightfirsttime.com/downloads/BriggsReportA4_FIN.pdf.

Burgess, R., Bishop, A., Lewis, M. and Hill, J., 2019. Models used for case-mix adjustment of patient reported outcome measures (PROMs) in musculoskeletal healthcare: A systematic review of the literature. *Physiotherapy*, *105*(2), pp.137-146.

Burgess, R., Mansell, G., Bishop, A., Lewis, M. and Hill, J., 2020a. Predictors of functional outcome in musculoskeletal healthcare: An umbrella review. *European Journal of Pain*, *24*(1), pp.51-70.

Burgess, R., Hall, J., Bishop, A., Lewis, M. and Hill, J., 2020b. Costing Methodology and Key Drivers of Health Care Costs Within Economic Analyses in Musculoskeletal Community and Primary Care Services: A Systematic Review of the Literature. *Journal of Primary Care & Community Health*, *11*, p.2150132719899763.

Burgess, R., Lewis, M., McRobert, C. and Hill, J.C., 2021a. Developing a Core Outcome Set for Community and Primary Care Musculoskeletal Services: A Consensus Approach. *Musculoskeletal Science and Practice*, p.102415.

Burgess, R., Lewis, M. and Hill, J.C., 2021b. Musculoskeletal case-mix adjustment in a UK primary/community care cohort: Testing musculoskeletal models to make recommendations in this setting. *Musculoskeletal Science and Practice*, *56*, p.102455.

Campbell, P., Hill, J.C., Protheroe, J., Afolabi, E.K., Lewis, M., Beardmore, R., Hay, E.M., Mallen, C.D., Bartlam, B., Saunders, B. and van der Windt, D.A., 2016. Keele Aches and Pains Study protocol: validity, acceptability, and feasibility of the Keele STarT MSK tool for subgrouping musculoskeletal patients in primary care. Journal of pain research, 9, p.807.

Chartered Society of Physiotherapy (CSP). 2020. National Evaluation of First Contact Practitioner (FCP) model of primary care. Available at:

https://www.csp.org.uk/system/files/documents/2020-11/final_fcp_phase_3_national_evaluation_report.pdf Accessed 13/10/21

Chartered Society of Physiotherapy (CSP) Website 2021a. https://www.csp.org.uk/news/2021-07-13-coming-soon-new-csp-msk-physiotherapy-service-standards-audit-tool Accessed 13/10/21

Chartered Society of Physiotherapy (CSP) Website 2021b. Image. Available at: https://www.csp.org.uk/professional-clinical/improvement-and-innovation/improving-and-transforming-your-service/intro Accessed 13/10/21

Coles, J. PROMs risk adjustment methodology guide for general surgery and orthopaedic procedures. Available at: https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2013/07/proms-ris-adj-meth-sur-orth.pdf Accessed 27/10/21

Department of Health., 2012. PROMs in England; the case-mix adjustment methodology. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/216507/dh_133449.pdf Accessed 27/10/21

Deutscher, D., Werneke, M.W., Hayes, D., Mioduski, J.E., Cook, K.F., Fritz, J.M., Woodhouse, L.J. and Stratford, P.W., 2018. Impact of Risk Adjustment on Provider Ranking for Patients With Low Back Pain Receiving Physical Therapy. *journal of orthopaedic & sports physical therapy*, *48*(8), pp.637-648.

Downie, W.W., Leatham, P.A., Rhind, V.M., Wright, V., Branco, J.A. and Anderson, J.A., 1978. Studies with pain rating scales. *Annals of the rheumatic diseases*, *37*(4), pp.378-381.

Doyle, C., Lennox, L. and Bell, D., 2013. A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. *BMJ open*, *3*(1).

Elwyn, G., Barr, P.J., Grande, S.W., Thompson, R., Walsh, T. and Ozanne, E.M., 2013. Developing CollaboRATE: a fast and frugal patient-reported measure of shared decision making in clinical encounters. *Patient education and counseling*, *93*(1), pp.102-107.

Fennelly, O., Blake, C., Desmeules, F., Stokes, D. and Cunningham, C., 2018. Patient-reported outcome measures in advanced musculoskeletal physiotherapy practice: a systematic review. *Musculoskeletal Care*, *16*(1), pp.188-208.

GIRFT, 2020. Getting It Right in Orthopaedics: Reflecting on success and reinforcing improvement. A follow-up on the GIRFT national specialty report on orthopaedics. Available at: https://gettingitrightfirsttime.co.uk/wp-content/uploads/2020/02/GIRFT-orthopaedics-follow-up-report-February-2020.pdf Accessed 11/10/21

GIRFT website, 2021. Available at: https://www.gettingitrightfirsttime.co.uk/girft-implementation/ Accessed 11/05/21.

Gutacker, N., Street, A., Gomes, M. and Bojke, C., 2015. Should English healthcare providers be penalised for failing to collect patient-reported outcome measures? A retrospective analysis. *Journal of the Royal Society of Medicine*, *108*(8), pp.304-316.

Hibbert, P., Saeed, F., Taylor, N., Clay-Williams, R., Winata, T., Clay, C., Hussein, W. and Braithwaite, J., 2020. Can benchmarking Australian hospitals for quality identify and improve high and low performers? Disseminating research findings for hospitals. *International Journal for Quality in Health Care*, *32*(Supplement_1), pp.84-88.

Hill, J.C., Whitehurst, D.G., Lewis, M., Bryan, S., Dunn, K.M., Foster, N.E., Konstantinou, K., Main, C.J., Mason, E., Somerville, S. and Sowden, G., 2011. Comparison of stratified primary care management for low back pain with current best practice (STarT Back): a randomised controlled trial. *The Lancet*, *378*(9802), pp.1560-1571.

Hill, J.C., Kang, S., Benedetto, E., Myers, H., Blackburn, S., Smith, S., Dunn, K.M., Hay, E., Rees, J., Beard, D. and Glyn-Jones, S., 2016. Development and initial cohort validation of the Arthritis Research UK Musculoskeletal Health Questionnaire (MSK-HQ) for use across musculoskeletal care pathways. *BMJ open*, *6*(8), p.e012331.

Hill, J.C., Garvin, S., Chen, Y., Cooper, V., Wathall, S., Saunders, B., Lewis, M., Protheroe, J., Chudyk, A., Dunn, K.M. and Hay, E., 2020. Stratified primary care versus non-stratified care for musculoskeletal pain: findings from the STarT MSK feasibility and pilot cluster randomized controlled trial. *BMC family practice*, *21*(1), pp.1-18.

Iezzoni, L.I., 2009. 3. 1 Risk adjustment for performance measurement. *Performance measurement for health system improvement: experiences, challenges and prospects*, *251*.

Jordan KP, Kadam UT, Hayward R, Porcheret M, Young C, Croft P. Annual consultation prevalence of regional musculoskeletal problems in primary care: an observational study. *BMC Musculoskelet Disord*.2010;11:144.

Kirkham, J.J., Davis, K., Altman, D.G., Blazeby, J.M., Clarke, M., Tunis, S. and Williamson, P.R., 2017. Core outcome Set-STAndards for development: the COS-STAD recommendations. *PLoS medicine*, *14*(11), p.e1002447.

Kyte, D.G., Calvert, M., Van der Wees, P.J., Ten Hove, R., Tolan, S. and Hill, J.C., 2015. An introduction to patient-reported outcome measures (PROMs) in physiotherapy. *Physiotherapy*, *101*(2), pp.119-125.

Lin, I., Wiles, L., Waller, R., Goucke, R., Nagree, Y., Gibberd, M., Straker, L., Maher, C.G. and O'Sullivan, P.P., 2020. What does best practice care for musculoskeletal pain look like? Eleven consistent recommendations from high-quality clinical practice guidelines: systematic review. British journal of sports medicine, 54(2), pp.79-86.

NHS Digital 2017. https://digital.nhs.uk/data-and-information/publications/statistical/patient-reported-outcome-measures-proms/hip-and-knee-replacement-procedures-april-2019-to-march-2020 Accessed 12/10/21

NHS Digital 2020. Provisional Patient Reported Outcome Measures (PROMs) for Hip and Knee Replacements in England April 2019 to March 2020: Data Quality Tables. https://files.digital.nhs.uk/F0/258782/PROMs%20Data%20Quality%20Tables%202019-20%20Provisional.xlsx Accessed 12/10/21

NHS Digital, 2021a. So, what is a Local Health and Care Record anyway? Available at: https://digital.nhs.uk/blog/transformation-blog/2019/so-what-is-a-local-health-and-care-record-anyway Accessed 12/10/21

NHS Digital, 2021b. Part of Digital inclusion for health and social care: What we mean by digital inclusion. Available at: https://digital.nhs.uk/about-nhs-digital/our-work/digital-inclusion/what-digital-inclusion-is Accessed 12/10/21

NHS England, 2013. Patient reported outcome measures in England: Update to reporting and case-mix adjusting hip and knee procedure data. Available at: https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2013/10/proms-meth-prim-revis.pdf Accessed 06/08/19

NHS England 2014. Five Year Forward View. Available at: https://www.england.nhs.uk/wp-content/uploads/2014/10/5yfv-web.pdf Accessed 06/08/19

NHS England, 2018. Bite-size guide to patient insight: The National Patient Reported Outcome Measures (PROMs) Programme. https://www.england.nhs.uk/wp-content/uploads/2018/08/proms-guide-aug-18-v3.pdf Accessed 11/10/21

NHS England, 2019. The Friends and Family Test Guide 2014. Available at: https://www.england.nhs.uk/wp-content/uploads/2019/09/using-the-fft-to-improve-patient-experience-guidance-v2.pdf Accessed 11/05/21

NHS England Website, 2020. Available at: https://www.nhs.uk/using-the-nhs/about-the-nhs/friends-and-family-test-fft/ Accessed 11/05/21

NHS England 2020a. Restoring primary and community musculoskeletal services: principles for integrated musculoskeletal service delivery. https://future.nhs.uk/gf2.ti/f/796802/79001541.4/PDF/-/C0832_MSK_guidance_211220qacc_V1.3_010321_NHSFUTURES_DRAFT.pdf Accessed 01/07/21

NHS England 2020b. The Improving Access to Psychological Therapies Manual. https://www.england.nhs.uk/wp-content/uploads/2020/05/iapt-manual-v4.pdf Accessed 12/10/21

NHS England and Ipsos MORI. GP Patient Survey: Questionnaire redevelopment. 2018. https://www.gp-patient.co.uk/Files/GPPS%20Y12%20Questionnaire%20redevelopment%20report%20v1%20PUBLIC.pdf Accessed 12/10/21

NHS Futures, 2021. Best MSK High Impact Strategy FINAL July 2021. Available at: https://future.nhs.uk/NationalMSKHealth/view?objectId=112835237 Accessed 12/10/21

NHS England and NHS Improvement, 2021. Model Health System. Available at: https://feedback.model.nhs.uk/knowledgebase/articles/1979268-funnel-plots Accessed 12/10/21

NICE, 2016. Low back pain and sciatica in over 16s: assessment and management. Available at: https://www.nice.org.uk/guidance/ng59 Accessed 23/11/21

NICE, 2021. Chronic pain (primary and secondary) in over 16s: assessment of all chronic pain and management of chronic primary pain: NICE guideline [NG193]Published: 07 April 2021. Available at: https://www.nice.org.uk/guidance/ng193 Accessed 23/11/21

Noben, C.Y., Evers, S.M., Nijhuis, F.J. and de Rijk, A.E., 2014. Quality appraisal of generic self-reported instruments measuring health-related productivity changes: a systematic review. *BMC Public Health*, *14*(1), pp.1-21.

Pfizer Limited. PP-PFE-GBR-2287 / Jan 2020. All rights reserved. Registered in England and Wales No. 526209. Registered office: Ramsgate Road, Sandwich, Kent CT13 9NJ. VAT No. GB201048427

Price, A.J., Ogollah, R., Kang, S., Hay, E., Barker, K.L., Benedetto, E., Smith, S., Smith, J., Galloway, J.B., Ellis, B. and Rees, J., 2019a. Determining responsiveness and meaningful changes for the Musculoskeletal Health Questionnaire (MSK-HQ) for use across musculoskeletal care pathways. *BMJ open*, *9*(10), p.e025357.

Price, C.M., C. de C. Williams, A., Smith, B.H. and Bottle, A., 2019b. Implementation of Patient-Reported Outcomes (PROMs) from specialist pain clinics in England and Wales: Experience from a nationwide study. *European Journal of Pain*, *23*(7), pp.1368-1377.

Porter, M.E., 2010. What is value in health care. *N Engl J Med*, *363*(26), pp.2477-2481.

Reilly, M.C., Zbrozek, A.S. and Dukes, E.M., 1993. The validity and reproducibility of a work productivity and activity impairment instrument. *Pharmacoeconomics*, *4*(5), pp.353-365.

Roberts, L., 2013. Improving quality, service delivery and patient experience in a musculoskeletal service. *Manual therapy*, *18*(1), pp.77-82.

Sajid, I., Parkunan, A., and Frost, K., 2020. Unintended Consequences: An Evaluation Quantifying Benefits, Iatrogenic Harms and Downstream Cascade Costs of Musculoskeletal Magnetic Resonance Imaging (MRI) in UK Primary Care (November 21, 2020). Available at SSRN: http://dx.doi.org/10.2139/ssrn.3735015 Accessed 23/11/21

Siemens, D.R., and Booth, C.M., 2017. Benchmarking our urological care: It's just the beginning. *Canadian Urological Association Journal,* 11(8), p.223.

Slater, H., and Briggs, A.M., 2017. Models of Care for musculoskeletal pain conditions: driving change to improve outcomes. *Pain Management.* 7(5), pp.351-357

Smith, M.J., and Choma, T.J., 2017. Patient satisfaction in musculoskeletal medicine. *Current reviews in musculoskeletal medicine*, *10*(2), pp.207-211.

Smith, P.C., Mossialos, E., Papanicolas, I. and Leatherman, S., 2008. Performance measurement for health system improvement: Experiences, Challenges, Prospects. Geneva, CH, World Health Organization.

Speerin, R., Needs, C., Chua, J., Woodhouse, L.J., Nordin, M., McGlasson, R. and Briggs, A.M., 2020. Implementing models of care for musculoskeletal conditions in health systems to support value-based care. *Best Practice & Research Clinical Rheumatology*, p.101548.

Stratford, P., Gill, C., Westaway, M. and Binkley, J., 1995. Assessing disability and change on individual patients: a report of a patient specific measure. *Physiotherapy canada*, *47*(4), pp.258-263.

Stynes S, Jordan KP, Hill JC, Wynne-Jones G, Cottrell E, Foster NE, Goodwin R, Bishop A. Evaluation of the First Contact Physiotherapy (FCP) model of primary care: Patient characteristics and outcomes, Physiotherapy, 2021.

Versus Arthritis (VA) 2019. State of musculoskeletal health 2019; Arthritis and other musculoskeletal conditions in numbers. Available at: https://www.versusarthritis.org/media/14594/state-of-musculoskeletal-health-2019.pdf Accessed 23/11/21

Wilkie, R. and Pransky, G., 2012. Improving work participation for adults with musculoskeletal conditions. *Best practice & research Clinical rheumatology*, *26*(5), pp.733-742.

Wilkie, R., Bjork, M., Costa-Black, K.M., Parker, M. and Pransky, G., 2020. Managing work participation for people with rheumatic and musculoskeletal diseases. *Best Practice & Research Clinical Rheumatology*, p.101517.

World Health Organization and International Telecommunication Union, 2020. Digital health platform handbook: building a digital information infrastructure (infostructure) for health. Geneva: Licence: CC BY-NC-SA 3.0 IGO. Available at: https://www.itu.int/en/myitu/Publications/2020/11/20/08/09/Building-a-Digital-Information-Infrastructure-for-Health Accessed 13/10/21

9 Appendices

Appendix 2-1: Case-mix adjustment systematic review: Search results

#	Database	Search term	Results
1	Medline	(physiotherap*).ti,ab	20170
2	Medline	("physical therap*").ti,ab	17332
3	Medline	(rheumatolog*).ti,ab	27857
4	Medline	(Orthopaedi*).ti,ab	32882
5	Medline	(Orthopedi*).ti,ab	35796
6	Medline	(Chiropract*).ti,ab	5047
7	Medline	(Osteopath*).ti,ab	4747
8	Medline	(rehabilitat*).ti,ab	132634
9	Medline	exp "PHYSICAL THERAPY MODALITIES"/	198318
10	Medline	REHABILITATION/	189286
11	Medline	exp "MUSCULOSKELETAL MANIPULATIONS"/	14266
12	Medline	(1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11)	539790
13	Medline	(low* ADJ back).ti,ab	27483
14	Medline	(cervical).ti,ab	184992
15	Medline	(spine).ti,ab	93636
16	Medline	(spinal).ti,ab	225643
17	Medline	(hip).ti,ab	110924
18	Medline	(knee).ti,ab	109237
19	Medline	(shoulder).ti,ab	52204

20	Medline	(Musculoskeletal).ti,ab	37510
21	Medline	SCIATICA/	4718
22	Medline	exp "BACK PAIN"/	32951
23	Medline	exp SPINE/	124297
24	Medline	HIP/	10837
25	Medline	exp "HIP JOINT"/	23967
26	Medline	exp "KNEE JOINT"/	50054
27	Medline	KNEE/	12391
28	Medline	SHOULDER/	11036
29	Medline	"SHOULDER JOINT"/	16237
30	Medline	SHOULDER/	11036
31	Medline	exp "MUSCULOSKELETAL PAIN"/	2627
32	Medline	exp "MUSCULOSKELETAL SYSTEM"/	1336749
33	Medline	exp "MUSCULOSKELETAL DISEASES"/	988774
34	Medline	(13 OR 14 OR 15 OR 16 OR 17 OR 18 OR 19 OR 20 OR 21 OR 22 OR 23 OR 24 OR 25 OR 26 OR 27 OR 28 OR 29 OR 30 OR 31 OR 32 OR 33)	2410888
35	Medline	(Baseline variabl*).ti,ab	36869
36	Medline	(Characteristic*).ti,ab	1080307
37	Medline	(demographic*).ti,ab	218641
38	Medline	(prognostic indicat*).ti,ab	40056
39	Medline	(predictor*).ti,ab	280753
40	Medline	exp "POPULATION CHARACTERISTICS"/	7252956

41	Medline	FORECASTING/	76553
42	Medline	(35 OR 36 OR 37 OR 38 OR 39 OR 40 OR 41)	8067345
43	Medline	(Patient Recorded Outcome Measure*).ti,ab	2458
44	Medline	(PROM).ti,ab	1835
45	Medline	(effectiveness).ti,ab	337302
46	Medline	("Change score*").ti,ab	2984
47	Medline	("Health gain*").ti,ab	1399
48	Medline	("Functional status*").ti,ab	20277
49	Medline	(43 OR 44 OR 45 OR 46 OR 47 OR 48)	364094
50	Medline	(case mix adjustment).ti,ab	946
51	Medline	(case-mix adjustment).ti,ab	934
52	Medline	("risk adjust*").ti,ab	6856
53	Medline	("regression analys*").ti,ab	200297
54	Medline	"RISK ADJUSTMENT"/	2594
55	Medline	exp "REGRESSION ANALYSIS"/	347144
56	Medline	(50 OR 51 OR 52 OR 53 OR 54 OR 55)	489201
57	Medline	(12 AND 34 AND 42 AND 49 AND 56)	401
58	Medline	(12 AND 34 AND 49 AND 56)	425
59	Medline	(mortalit*).ti,ab	578260
60	Medline	(random* ADJ2 trial*).ti,ab	253649
61	Medline	(59 OR 60)	804076
62	Medline	57 NOT 61	293

63	Medline	58 NOT 61	313
64	Medline	62 [DT 1992-2017] [Human age groups Young adult OR Adult OR Middle Aged OR Aged OR Aged,80 and over] [Languages English]	246
65	Medline	63 [DT 1992-2017] [Human age groups Young adult OR Adult OR Middle Aged OR Aged OR Aged,80 and over] [Languages English]	252
66	CINAHL	(physiotherap*).ti,ab	11682
67	CINAHL	("physical therap*").ti,ab	11632
68	CINAHL	(rheumatolog*).ti,ab	3828
69	CINAHL	(Orthopaedi*).ti,ab	7483
70	CINAHL	(Orthopedi*).ti,ab	6425
71	CINAHL	(Chiropract*).ti,ab	9111
72	CINAHL	(Osteopath*).ti,ab	1984
73	CINAHL	(rehabilitat*).ti,ab	52640
74	CINAHL	exp "PHYSICAL THERAPY"/	78787
75	CINAHL	REHABILITATION/	11437
76	CINAHL	"MANIPULATION, OSTEOPATHIC"/	459
77	CINAHL	"MANIPULATION, CHIROPRACTIC"/	3131
78	CINAHL	"MANIPULATION, ORTHOPEDIC"/	1558
79	CINAHL	(66 OR 67 OR 68 OR 69 OR 70 OR 71 OR 72 OR 73 OR 74 OR 75 OR 76 OR 77 OR 78)	181910
80	CINAHL	(low* ADJ back).ti,ab	10757
81	CINAHL	(cervical).ti,ab	19261
82	CINAHL	(spine).ti,ab	14196

83	CINAHL	(spinal).ti,ab	26868
84	CINAHL	(hip).ti,ab	20897
85	CINAHL	(knee).ti,ab	23153
86	CINAHL	(shoulder).ti,ab	11920
87	CINAHL	(Musculoskeletal).ti,ab	11078
88	CINAHL	SCIATICA/	721
89	CINAHL	exp "BACK PAIN"/	17173
90	CINAHL	exp SPINE/	18952
91	CINAHL	"HIP JOINT"/	2868
92	CINAHL	HIP/	3762
93	CINAHL	exp "KNEE JOINT"/	5979
94	CINAHL	KNEE/	5803
95	CINAHL	exp "SHOULDER JOINT"/	2330
96	CINAHL	SHOULDER/	3598
97	CINAHL	exp "MUSCULOSKELETAL DISEASES"/	119977
98	CINAHL	exp "MUSCULOSKELETAL SYSTEM"/	105978
99	CINAHL	(80 OR 81 OR 82 OR 83 OR 84 OR 85 OR 86 OR 87 OR 88 OR 89 OR 90 OR 91 OR 92 OR 93 OR 94 OR 95 OR 96 OR 97 OR 98)	279553
100	CINAHL	(Patient Recorded Outcome Measure*).ti,ab	11
101	CINAHL	(PROM).ti,ab	398
102	CINAHL	(effectiveness).ti,ab	63815
103	CINAHL	("Change score*").ti,ab	943

104	CINAHL	("Health gain*").ti,ab	421
105	CINAHL	("Functional status*").ti,ab	5417
106	CINAHL	(100 OR 101 OR 102 OR 103 OR 104 OR 105)	76343
107	CINAHL	(case mix adjustment).ti,ab	132
108	CINAHL	(case-mix adjustment).ti,ab	132
109	CINAHL	("risk adjust*").ti,ab	1768
110	CINAHL	("regression analys*").ti,ab	31494
111	CINAHL	exp REGRESSION/	156781
112	CINAHL	(107 OR 108 OR 109 OR 110 OR 111)	170902
113	CINAHL	(79 AND 99 AND 106 AND 112)	464
114	CINAHL	(mortalit*).ti,ab	72668
115	CINAHL	(random* ADJ2 trial*).ti,ab	63600
116	CINAHL	(114 OR 115)	130287
117	CINAHL	113 NOT 116	347
118	CINAHL	117 [DT 1992-2017] [Human age groups All Adult] [Languages eng]	175
119	EMBASE	(physiotherap*).ti,ab	34117
120	EMBASE	("physical therap*").ti,ab	25477
121	EMBASE	(rheumatolog*).ti,ab	59881
122	EMBASE	(Orthopaedi*).ti,ab	48120
123	EMBASE	(Orthopedi*).ti,ab	45062
124	EMBASE	(Chiropract*).ti,ab	4949
125	EMBASE	(Osteopath*).ti,ab	5831

126	EMBASE	(rehabilitat*).ti,ab	184847
127	EMBASE	exp PHYSIOTHERAPY/	79479
128	EMBASE	REHABILITATION/	122759
129	EMBASE	exp "MANIPULATIVE MEDICINE"/	32718
130	EMBASE	(119 OR 120 OR 121 OR 122 OR 123 OR 124 OR 125 OR 126 OR 127 OR 128 OR 129)	459873
131	EMBASE	(low* ADJ back).ti,ab	37015
132	EMBASE	(cervical).ti,ab	237567
133	EMBASE	(spine).ti,ab	127837
134	EMBASE	(spinal).ti,ab	285153
135	EMBASE	(hip).ti,ab	144724
136	EMBASE	(knee).ti,ab	140456
137	EMBASE	(shoulder).ti,ab	64116
138	EMBASE	(Musculoskeletal).ti,ab	49420
139	EMBASE	exp "MUSCULOSKELETAL PAIN"/	123194
140	EMBASE	exp BACKACHE/	91685
141	EMBASE	exp SPINE/	189226
142	EMBASE	HIP/	72469
143	EMBASE	KNEE/	89997
144	EMBASE	SHOULDER/	42907
145	EMBASE	exp "MUSCULOSKELETAL PAIN"/	123194
146	EMBASE	exp "MUSCULOSKELETAL SYSTEM"/	1828968
147	EMBASE	exp "MUSCULOSKELETAL DISEASE"/	2007170

148	EMBASE	(131 OR 132 OR 133 OR 134 OR 135 OR 136 OR 137 OR 138 OR 139 OR 140 OR 141 OR 142 OR 143 OR 144 OR 145 OR 146 OR 147)	3462628
149	EMBASE	(Patient Recorded Outcome Measure*).ti,ab	12
150	EMBASE	(PROM).ti,ab	2836
151	EMBASE	(effectiveness).ti,ab	449707
152	EMBASE	("Change score*").ti,ab	4356
153	EMBASE	("Health gain*").ti,ab	1833
154	EMBASE	("Functional status*").ti,ab	29474
155	EMBASE	(149 OR 150 OR 151 OR 152 OR 153 OR 154)	485494
156	EMBASE	(case mix adjustment).ti,ab	470
157	EMBASE	(case-mix adjustment).ti,ab	470
158	EMBASE	("risk adjust*").ti,ab	9695
159	EMBASE	("regression analys*").ti,ab	284183
160	EMBASE	exp "REGRESSION ANALYSIS"/	538272
161	EMBASE	(156 OR 157 OR 158 OR 159 OR 160)	628718
162	EMBASE	(130 AND 148 AND 155 AND 161)	583
163	EMBASE	(mortalit*).ti,ab	836322
164	EMBASE	(random* ADJ2 trial*).ti,ab	307228
165	EMBASE	(163 OR 164)	1106759
166	EMBASE	162 NOT 165	460
167	EMBASE	166 [DT 1992-2017] [English language] [Human age groups Adult 18 to 64 years OR Aged 65+ years]	287
168	AMED	(physiotherap*).ti,ab	6387

169	AMED	("physical therap*").ti,ab	4896
170	AMED	(rheumatolog*).ti,ab	605
171	AMED	(Orthopaedi*).ti,ab	1610
172	AMED	(Orthopedi*).ti,ab	1092
173	AMED	(Chiropract*).ti,ab	4156
174	AMED	(Osteopath*).ti,ab	1484
175	AMED	(rehabilitat*).ti,ab	24355
176	AMED	REHABILITATION/	51349
177	AMED	exp "PHYSICAL THERAPY MODALITIES"/	23837
178	AMED	exp "MUSCULOSKELETAL MANIPULATIONS"/	5235
179	AMED	(168 OR 169 OR 170 OR 171 OR 172 OR 173 OR 174 OR 175 OR 176 OR 177 OR 178)	88031
180	AMED	(low* ADJ back).ti,ab	5114
181	AMED	(cervical).ti,ab	3243
182	AMED	(spine).ti,ab	4188
183	AMED	(spinal).ti,ab	8493
184	AMED	(hip).ti,ab	5127
185	AMED	(knee).ti,ab	9261
186	AMED	(shoulder).ti,ab	4018
187	AMED	(Musculoskeletal).ti,ab	3476
188	AMED	exp BACKACHE/	5999
189	AMED	exp SPINE/	5141
190	AMED	exp HIP/	796

191	AMED	"SHOULDER JOINT"/	989
192	AMED	"KNEE JOINT"/	3393
193	AMED	"HIP JOINT"/	1056
194	AMED	KNEE/	1498
195	AMED	SHOULDER/	1250
196	AMED	"MUSCULOSKELETAL PAIN"/	139
197	AMED	exp "MUSCULOSKELETAL DISEASE"/	27250
198	AMED	exp "MUSCULOSKELETAL SYSTEM"/	28665
199	AMED	(180 OR 181 OR 182 OR 183 OR 184 OR 185 OR 186 OR 187 OR 188 OR 189 OR 190 OR 191 OR 192 OR 193 OR 194 OR 195 OR 196 OR 197 OR 198)	66903
200	AMED	(Patient Recorded Outcome Measure*).ti,ab	0
201	AMED	(PROM).ti,ab	59
202	AMED	(effectiveness).ti,ab	8187
203	AMED	("Change score*").ti,ab	358
204	AMED	("Health gain*").ti,ab	31
205	AMED	("Functional status*").ti,ab	1337
206	AMED	(200 OR 201 OR 202 OR 203 OR 204 OR 205)	9794
207	AMED	(case mix adjustment).ti,ab	7
208	AMED	(case-mix adjustment).ti,ab	7
209	AMED	("risk adjust*").ti,ab	37
210	AMED	("regression analys*").ti,ab	2635
211	AMED	(207 OR 208 OR 209 OR 210)	2677

212	AMED	(179 AND 199 AND 206 AND 211)	47
213	AMED	(mortalit*).ti,ab	2147
214	AMED	(random* ADJ2 trial*).ti,ab	6722
215	AMED	(213 OR 214)	8720
216	AMED	212 NOT 215	40
217	AMED	216 [DT 1992-2017] [Languages English]	39
218	HMIC	(physiotherap*).ti,ab	1282
219	HMIC	("physical therap*").ti,ab	104
220	HMIC	(rheumatolog*).ti,ab	202
221	HMIC	(Orthopaedi*).ti,ab	1008
222	HMIC	(Orthopedi*).ti,ab	26
223	HMIC	(Chiropract*).ti,ab	94
224	HMIC	(Osteopath*).ti,ab	91
225	HMIC	(rehabilitat*).ti,ab	3503
226	HMIC	exp PHYSIOTHERAPY/	879
227	HMIC	"MEDICAL REHABILITATION"/	77
228	HMIC	REHABILITATION/	1707
229	HMIC	(218 OR 219 OR 220 OR 221 OR 222 OR 223 OR 224 OR 225 OR 226 OR 227 OR 228)	6746
230	HMIC	(low* ADJ back).ti,ab	249
231	HMIC	(cervical).ti,ab	1387
232	HMIC	(spine).ti,ab	150
233	HMIC	(spinal).ti,ab	282

234	HMIC	(hip).ti,ab	926
235	HMIC	(knee).ti,ab	323
236	HMIC	(shoulder).ti,ab	132
237	HMIC	(Musculoskeletal).ti,ab	477
238	HMIC	exp "BACK PAIN"/	395
239	HMIC	exp "SPINAL COLUMN"/	35
240	HMIC	"HIP BONES"/	88
241	HMIC	"HIP JOINTS"/	152
242	HMIC	"KNEE JOINTS"/	21
243	HMIC	KNEES/	43
244	HMIC	"SHOULDER BONES"/	0
245	HMIC	SHOULDERS/	21
246	HMIC	exp "MUSCULOSKELETAL SYSTEM"/	529
247	HMIC	exp "MUSCULOSKELETAL SYSTEM DISEASES"/	1295
248	HMIC	(230 OR 231 OR 232 OR 233 OR 234 OR 235 OR 236 OR 237 OR 238 OR 239 OR 240 OR 241 OR 242 OR 243 OR 244 OR 245 OR 246 OR 247)	4814
249	HMIC	(Patient Recorded Outcome Measure*).ti,ab	0
250	HMIC	(PROM).ti,ab	24
251	HMIC	(effectiveness).ti,ab	13455
252	HMIC	("Change score*").ti,ab	30
253	HMIC	("Health gain*").ti,ab	609
254	HMIC	("Functional status*").ti,ab	222

255	HMIC	(249 OR 250 OR 251 OR 252 OR 253 OR 254)	14151
256	HMIC	(case mix adjustment).ti,ab	42
257	HMIC	(case-mix adjustment).ti,ab	42
258	HMIC	("risk adjust*").ti,ab	304
259	HMIC	("regression analys*").ti,ab	2198
260	HMIC	"REGRESSION ANALYSIS"/	57
261	HMIC	(256 OR 257 OR 258 OR 259 OR 260)	2549
262	HMIC	(229 AND 248 AND 255 AND 261)	2

Appendix 3-1: Predictors of functional outcome umbrella review: Search results.

#	Database	Search term	Results
1	Medline	(predic*).ti,ab	1287666
2	Medline	(prognos*).ti,ab	487751
3	Medline	PROGNOSIS/	435890
4	Medline	(1 OR 2 OR 3)	1844789
5	Medline	(Outcom*).ti,ab	1344265
6	Medline	(Recovery).ti,ab	378897
7	Medline	(Function*).ti,ab	3073833
8	Medline	exp "PATIENT OUTCOME ASSESSMENT"/	4518
9	Medline	"RECOVERY OF FUNCTION"/	42074
10	Medline	(5 OR 6 OR 7 OR 8 OR 9)	4454961
11	Medline	(4 AND 10)	595671
12	Medline	(musculoskeletal).ti	11557
13	Medline	(Low back).ti	12963
14	Medline	(Low* back).ti	13585
15	Medline	(Neck).ti	64516
16	Medline	Spine OR (Spinal).ti	261448
17	Medline	Hip OR (Hips).ti	139665
18	Medline	Knee OR (knees).ti	148208
19	Medline	Shoulder OR (shoulders).ti	70578

20	Medline	(12 OR 13 OR 14 OR 15 OR 16 OR 17 OR 18 OR 19)	657309
21	Medline	(11 AND 20)	23074
22	Medline	("systematic revie*").ti	78410
23	Medline	(21 AND 22)	438
24	Medline	23 [DT 2012-2017] [Languages English]	328
25	EMBASE	(predic*).ti,ab	1733429
26	EMBASE	(prognos*).ti,ab	736747
27	EMBASE	PROGNOSIS/ OR "PROGNOSTIC ASSESSMENT"/	534442
28	EMBASE	(25 OR 26 OR 27)	2398276
29	EMBASE	(Outcom*).ti,ab	1981947
30	EMBASE	(Recovery).ti,ab	501093
31	EMBASE	(Function*).ti,ab	3912509
34	EMBASE	exp "OUTCOME ASSESSMENT"/	418191
35	EMBASE	(29 OR 30 OR 31 OR 34)	6008440
36	EMBASE	(28 AND 35)	853037
37	EMBASE	(musculoskeletal).ti	14737
38	EMBASE	(Low back).ti	16379
39	EMBASE	(Low* back).ti	17103
40	EMBASE	(Neck).ti	79115
41	EMBASE	Spine OR (Spinal).ti	349259

42	EMBASE	Hip OR (Hips).ti	193047
43	EMBASE	Knee OR (knees).ti	196147
44	EMBASE	Shoulder OR (shoulders).ti	89013
45	EMBASE	(37 OR 38 OR 39 OR 40 OR 41 OR 42 OR 43 OR 44)	835010
46	EMBASE	(36 AND 45)	30755
47	EMBASE	("systematic revie*").ti	97247
48	EMBASE	(46 AND 47)	429
49	EMBASE	48 [DT 2012-2017] [English language]	328
50	EMBASE	48 [DT 2012-2017] [English language] [Human age groups Adult 18 to 64 years OR Aged 65+ years]	47
51	Medline	23 [DT 2012-2017] [Human age groups Young adult OR Adult OR Middle Aged OR Aged OR Aged,80 and over] [Languages English]	51
52	CINAHL	(predic*).ti,ab	152815
53	CINAHL	(prognos*).ti,ab	38378
54	CINAHL	PROGNOSIS/	29150
55	CINAHL	(52 OR 53 OR 54)	195175
56	CINAHL	(Outcom*).ti,ab	284195
57	CINAHL	(Recovery).ti,ab	38201
58	CINAHL	(Function*).ti,ab	199573
59	CINAHL	"PATIENT-REPORTED OUTCOMES"/	181
60	CINAHL	"OUTCOME ASSESSMENT"/	21875

61	CINAHL	(56 OR 57 OR 58 OR 59 OR 60)	468571
62	CINAHL	(55 AND 61)	72263
63	CINAHL	(musculoskeletal).ti	5164
64	CINAHL	(Low back).ti	7263
65	CINAHL	(Low* back).ti	7582
66	CINAHL	(Neck).ti	10977
67	CINAHL	Spine OR (Spinal).ti	56191
68	CINAHL	Hip OR (Hips).ti	52645
69	CINAHL	Knee OR (knees).ti	55257
70	CINAHL	Shoulder OR (shoulders).ti	37570
71	CINAHL	(63 OR 64 OR 65 OR 66 OR 67 OR 68 OR 69 OR 70)	181915
72	CINAHL	(62 AND 71)	6024
73	CINAHL	("systematic revie*").ti	24551
74	CINAHL	(72 AND 73)	116
75	CINAHL	74 [DT 2012-2017] [Languages eng]	73
76	CINAHL	74 [Human age groups All Adult]	17
100	CINAHL	74 [DT 2012-2017] [Human age groups All Adult] [Languages eng]	10
77	AMED	(predic*).ti,ab	11232
78	AMED	(prognos*).ti,ab	2383
79	AMED	PROGNOSIS/	1966
80	AMED	(77 OR 78 OR 79)	14080

81	AMED	(Outcom*).ti,ab	27577
82	AMED	(Recovery).ti,ab	6172
83	AMED	(Function*).ti,ab	33975
84	AMED	"TREATMENT OUTCOME"/	16607
85	AMED	(81 OR 82 OR 83 OR 84)	64678
86	AMED	(80 AND 85)	6662
87	AMED	(musculoskeletal).ti	1272
88	AMED	(Low back).ti	3432
89	AMED	(Low* back).ti	3538
90	AMED	(Neck).ti	1622
91	AMED	Spine OR (Spinal).ti	9972
92	AMED	Hip OR (Hips).ti	5668
93	AMED	Knee OR (knees).ti	10801
94	AMED	Shoulder OR (shoulders).ti	4700
95	AMED	(87 OR 88 OR 89 OR 90 OR 91 OR 92 OR 93 OR 94)	33154
96	AMED	(86 AND 95)	1390
97	AMED	("systematic revie*").ti	2440
98	AMED	(96 AND 97)	17
99	AMED	98 [DT 2012-2017] [Languages English]	4

#	Database	Search term	Results
1	Medline	(Economic OR "Cost consequence" OR "Cost utility" OR "Cost benefit" OR "Cost effectiveness" OR "Cost minimisation").ti	55457
2	Medline	(Economic OR "Cost consequence" OR "Cost utility" OR "Cost benefit" OR "Cost effectiveness" OR "Cost minimization").ti	55679
3	Medline	(1 OR 2)	55751
4	Medline	(analysis OR evaluation).ti	1189179
5	Medline	(3 AND 4)	14380
6	Medline	(Musculoskeletal OR back* OR Neck* OR Spinal OR Spine OR Knee* OR Shoulder*).ti	360132
7	Medline	(5 AND 6)	302
8	Medline	exp "MUSCULOSKELETAL PAIN"/ OR exp "MUSCULOSKELETAL SYSTEM"/	1412585
9	Medline	(5 AND 8)	174
10	Medline	(6 OR 8)	1663038
11	Medline	(5 AND 10)	425
12	Medline	11 [DT 2008-2018] [Languages English]	301
13	Medline	7 [DT 2008-2018] [Languages English]	220
14	CINAHL	(Economic OR "Cost consequence" OR "Cost utility" OR "Cost benefit" OR "Cost effectiveness" OR "Cost minimisation" OR "Cost minimization").ti	21959

15	CINAHL	(analysis OR evaluation).ti	207080
16	CINAHL	(Musculoskeletal OR back* OR Neck* OR Spinal OR Spine OR Knee* OR Shoulder*).ti	135334
17	CINAHL	(14 AND 15 AND 16)	205
18	CINAHL	17 [DT 2008-2018] [Languages eng]	164
19	CINAHL	exp "MUSCULOSKELETAL SYSTEM"/	202693
20	AMED	(Economic OR "Cost consequence" OR "Cost utility" OR "Cost benefit" OR "Cost effectiveness" OR "Cost minimisation" OR "Cost minimization").ti	576
21	AMED	(analysis OR evaluation).ti	13160
22	AMED	(Musculoskeletal OR back* OR Neck* OR Spinal OR Spine OR Knee* OR Shoulder*).ti	22978
23	AMED	(20 AND 21 AND 22)	31
24	AMED	23 [DT 2008-2018] [Languages English]	11
25	BNI	(Economic OR "Cost consequence" OR "Cost utility" OR "Cost benefit" OR "Cost effectiveness" OR "Cost minimisation" OR "Cost minimization").ti	2213
26	BNI	(analysis OR evaluation).ti	21248
27	BNI	(Musculoskeletal OR back* OR Neck* OR Spinal OR Spine OR Knee* OR Shoulder*).ti	8613
28	BNI	(25 AND 26 AND 27)	14
29	BNI	28 [DT 2008-2018]	9

30	EMBASE	(Economic OR "Cost consequence" OR "Cost utility" OR "Cost benefit" OR "Cost effectiveness" OR "Cost minimisation" OR "Cost minimization").ti	72933
31	EMBASE	(analysis OR evaluation).ti	1429402
32	EMBASE	(Musculoskeletal OR back* OR Neck* OR Spinal OR Spine OR Knee* OR Shoulder*).ti	430581
33	EMBASE	(30 AND 31 AND 32)	449
34	EMBASE	33 [DT 2008-2018] [Languages English]	339
35	HBE	(Economic OR "Cost consequence" OR "Cost utility" OR "Cost benefit" OR "Cost effectiveness" OR "Cost minimisation" OR "Cost minimization").ti	16111
36	HBE	(analysis OR evaluation).ti	18230
37	HBE	(Musculoskeletal OR back* OR Neck* OR Spinal OR Spine OR Knee* OR Shoulder*).ti	29173
38	HBE	(35 AND 36 AND 37)	5
39	HMIC	(Economic OR "Cost consequence" OR "Cost utility" OR "Cost benefit" OR "Cost effectiveness" OR "Cost minimisation" OR "Cost minimization").ti	4119
40	HMIC	(analysis OR evaluation).ti	13345
41	HMIC	(Musculoskeletal OR back* OR Neck* OR Spinal OR Spine OR Knee* OR Shoulder*).ti	2295
42	HMIC	(39 AND 40 AND 41)	17

43	HMIC	42 [DT 2008-2018] [Languages English]	10

Appendix 4-2: Highest mean costs per patient (with costs converted to GDP 2018)

Mean cost per patient converted/inflated to GDP 2018

	Primary Care & Community						Secondary Care			Imaging/Investigations		Medication	Private	
	GP*	Physio-therapy*	Manual Therapy*	Psychologist*	Complementary Medicine/ET/Massage*	Other HCP*	Outpatient visit*	Hospital stay day case*	Hospital admission*	Xray^	MRI^	Prescription medicines^	Private other HCP*	Equipment/aids^
Author (first)														
Barton 2009														
Bosmans 2011	10.79	24.30	10.94			0.00	2.99			1.12	3.18	8.92		
Bultmann 2009														
Chuang 2012														
Essex 2017	135.00	44.80			5.42	0.00	91.97	68.98	62.31			35.46		
Haines 2017														
Henchoz 2010	*No cost year stated - conversion not possible*													
Hollinghurst 2008	14.14					0.45						12.59	217.49	
Hollinghurst 2013	22.32	81.03				0.03	35.72		59.29			12.83		
Jenson 2017	277.84	421.30				39.04	357.74		238.79			13.12	559.31	
Jowett 2013						136.07						2.95	18.79	
Lamb 2010														

Lambeek 2010																		
Marra 2014	15.18				16.52		7.35			14.02	26.58		8.23	58.64				
McKenna 2009	106.41	33.33				12.82	67.95						98.71	123.07				
Pinto 2013																		
Saha 2018	249.66	286.00		104.34		63.34	64.32											
Smeets 2009	45.78	180.17		21.99		2.59	92.79		32.02			54.04						
van de Roer 2008	30.15	50.99	6.80		4.92	0.00	11.95			13.32		11.59						
Vermeulen 2013	22.92	245.49	84.15		15.96	35.30	122.74		85.99	37.49	50.20	214.46						
Werner 2016																		
Whitehurst 2015	48.24	40.38					41.19		19.52	4.36	24.39	22.68	65.06					

**Non-Tradeable goods - Inflated using local inflators, and converted to GBP using current exchange rate*

^ Tradeable goods - Converted to GBP using exchange rate in the price year, and inflated using UK inflators.

Top 2 (1st/2nd) mean costs within studies

Next 2 (3rd/4th) top mean costs within studies

Cost drivers only included within this table if they were within the top 4 mean costs within at least 1 study

Mean costs reported for usual care/control/reference group

HCP; Health Care Professional, ET; Exercise Therapy

Note: Costs converted by Health Economist James Hall who also acted as second reviewer for the review.

Appendix 5-1: MSK standardised dataset: Consensus survey protocol

Project Overview

The purpose of this project is to develop consensus on a minimum standardised dataset (Core Outcome Set (COS)) for use in musculoskeletal (MSK) community and primary care services. This agreed minimum dataset would include patient characteristics for descriptive analysis of MSK data, variables explaining complexity of the patient population in order to make fair comparisons, patient reported outcome measures (PROMs) to measure effectiveness of treatment, patient reported experience measures (PREMs) to measure patient experience, and any other agreed useful tools. The purpose of widespread data collection is for service evaluation, audit, and benchmarking services, identifying best practice and underperforming services, alongside providing the ability for structured and tailored quality improvement. The standardised dataset however needs to be kept to a minimum to ensure feasibility of collection across MSK services nationally and therefore needs to undergo a national consensus approach to reach agreement on which metrics are essential to collect, which are beneficial but not seen to be essential, and which are not useful for widespread collection. This agreed standardised dataset could then be used to develop a national MSK audit focused to community and primary care services which is currently lacking, to help transform services for the future.

Primary Objective

The primary aim is to reach consensus on the minimum dataset (core outcome set) that should be collected across MSK services in order to enable effective service evaluation and benchmarking (allowing for case-mix adjustment to ensure fair comparisons can be made, and including optimum PROMs/metrics to measure effectiveness and allow for quality improvement initiatives/evaluation within community and primary care settings). This dataset needs to be considered feasible and appropriate for collection by clinicians in clinical systems, and feasible

and appropriate to patients who will provide the majority of data in the form of questionnaires to be collected and collated by individual MSK services.

Secondary Objective

The secondary objective is to aid development of methodology for a national audit of community and primary care MSK services.

Participants

Healthcare Professionals: MSK clinicians, managers, commissioners/stakeholders in the UK who are interested and/or have expertise in the area of MSK data and MSK Community/Primary Care practice will be invited to participate. Professionals will therefore be adults of working age with a high educational ability.

Patients/Service Users: Inclusion Criteria; Patients/MSK Service Users who are 18 years or over and who have accessed MSK community/primary care services within the last 12 months for their MSK problem.

Recruitment

We intend to recruit a minimum of 100 clinicians/managers with the aim of having representation from across the UK and across professional groups, targeting those likely to have knowledge and expertise in MSK data. We also aim to recruit a minimum of 25 patients in order to gain essential feedback on feasibility and acceptability of a proposed patient questionnaire to patients using MSK services.

Email invitation: to professional networks including the consultant physiotherapist network, health professionals who have registered to hold a licence with Oxford Innovations for the MSK-HQ or who have directly contacted the research team with an interest in this area, the Versus Arthritis MSK Champions group and the Versus Arthritis Data Group (see Appendix 5-1-1 for draft email).

Communications: the survey invitation and link will also appear in the Versus Arthritis e-bulletin Network News in September (distribution n=7000), and will be posted on social media (Twitter/Linked In) with tweets/posts from Keele University's School of Primary, Community and Social Care, Dr Jonathan Hill and Roanna Burgess alongside Versus Arthritis' social media communications to gain further interest from the MSK stakeholder community including patient groups. Please see Appendix 5-1-2 for Recruitment and Survey Flow Diagram.

Consent

Consent will be implied through participants clicking on the link to the survey and after viewing the introduction videos continuing to the next page to start the survey questions. Any incomplete surveys will be excluded.

Survey

The survey should not take more than 30 minutes for health professionals or patients to complete. Respondents will be signposted to the appropriate questions (healthcare professional or patient/service user) dependent on their responses to initial questions. Please see Appendix 5-1-2 for Recruitment and Survey Flow Diagram and Appendix 3 and 4 for healthcare professional and patient/service user survey content respectively.

Confidentiality

The survey will be fully anonymised so that no identifiable information is collected or retained within the Health Survey software.

Right to Withdraw

Respondents have the right to withdraw at any time and if they do so without completing the survey then their results will be excluded from the analysis.

Analysis Plan

Data Analysis: questions and responses to the survey will be numbered and coded within the Keele Health Survey software, and on completion of data collection the anonymised data sitting within Health Survey will be exported to MS Excel for collation.

Determining scoring system/definition of consensus determined a priori: Consensus will be defined as agreement from at least 70% of the voting participants/stakeholders as supported by the COS-STAD development group (Kirkham et al, 2017). This will be agreement between clinicians/health professionals on core metrics that are classed as 'extremely' or 'very important' for inclusion in a standardised MSK dataset.

Criteria for inclusion/exclusion of outcomes/metrics: Inclusion/exclusion of metrics/outcome tools will be decided through analysis of how participants have rated metrics (extremely important, very important, moderately important, neutral, slightly important, low importance, not at all important) (Vagias, 2006), and the 70% pre-determined agreement level, alongside supporting questions on feasibility, and on patient feedback and comments on the proposed patient questionnaire.

Unambiguous language used to describe outcomes/metrics considered for inclusion: Patient and public involvement (PPI) will be used to ensure that the finalised list of metrics and wording around included questions makes sense to the target/user population. Any questions highlighted as unclear by patients/service users will be amended.

Collation of results: Results will be written up into a report and submitted for publication with the final agreed recommended standardised dataset also being uploaded to the Keele University website (alongside a data dictionary and a proposed patient questionnaire) allowing for interested clinicians to download and implement locally and allow for interested patients to view in detail the questions included in the final recommendation.

Data Storage: All of the anonymised online feedback held on the Keele Health Survey Software will be deleted following completion of the analysis and write-up of the survey results.

References:

Kirkham, J.J., Davis, K., Altman, D.G., Blazeby, J.M., Clarke, M., Tunis, S. and Williamson, P.R., 2017. Core outcome Set-STAndards for development: the COS-STAD recommendations. *PLoS medicine*, *14*(11), p.e1002447.

Vagias, Wade M. (2006). *"Likert-type scale response anchors.* Clemson International Institute for Tourism & Research Development, Department of Parks, Recreation and Tourism Management. Clemson University

Appendix 5-1-1: Email/Tweet/Addition to VA Network News E-Bulletin/Addition to the Keele MSK Tracker Webpage

We need you to help us improve NHS Services for the future! Given recent changes to NHS services in response to the COVID-19 pandemic it is more important than ever to evaluate the quality of care provided and to capture any changes in quality associated with new ways of working.

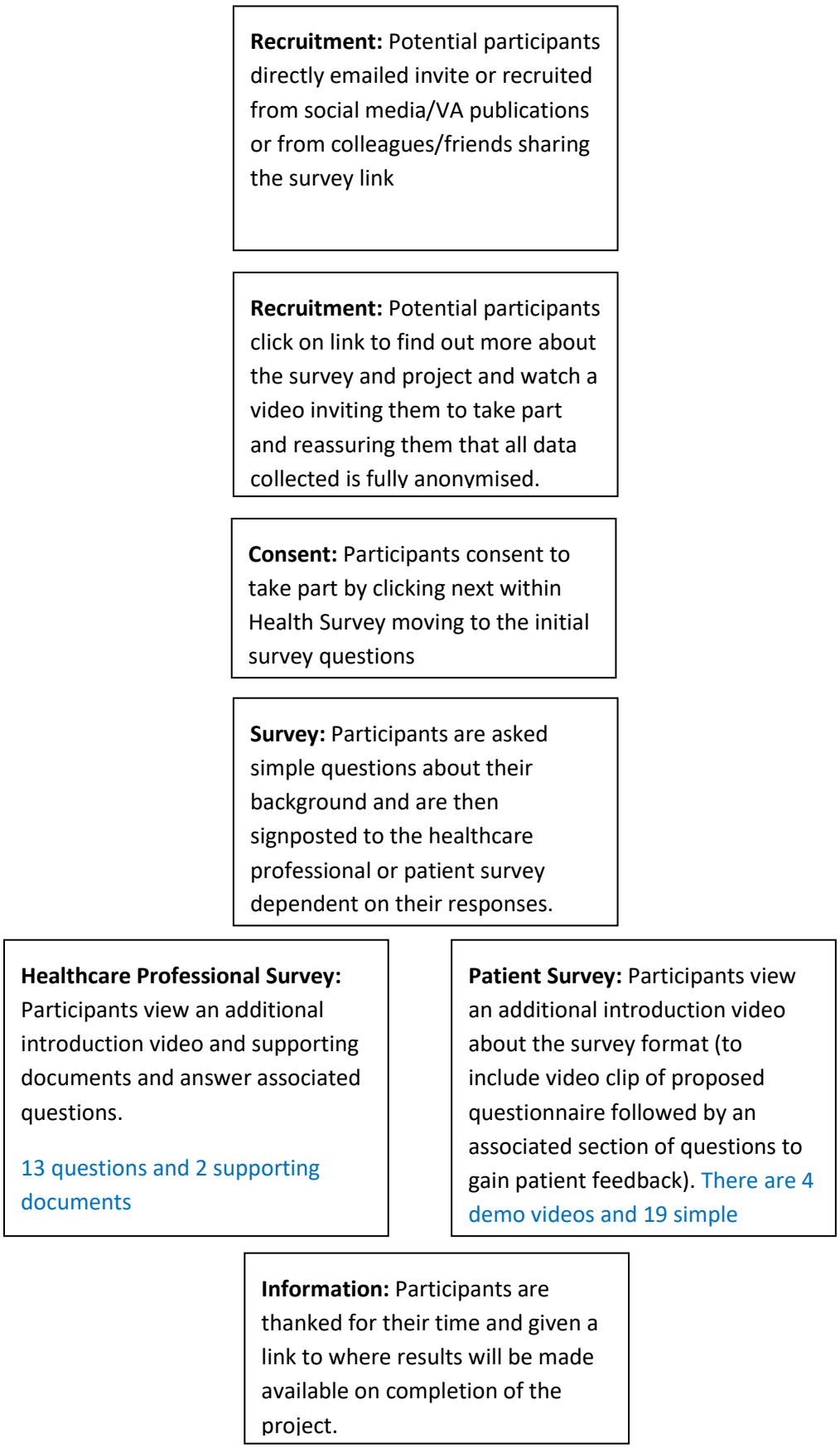
Our musculoskeletal (MSK) research team at Keele University is asking people who work in/manage or have recent experience of receiving MSK care within a community or primary care MSK setting (e.g. GP Practice, Physiotherapy Clinic) to complete this short survey to help us finalise a short-list for a standardised dataset for routine MSK service evaluation and benchmarking in this setting.

The survey is fully anonymised. Please click the link below to take part.

We would also be grateful if you could forward this email to any other Clinicians/Managers with expertise in this area or to friends and family who have recently received MSK care in this setting that you feel would be keen to participate.

https://healthsurvey.hfac.keele.ac.uk/index.php/375144?lang=en

Appendix 5-1-2: Recruitment/Survey Flow



Appendix 5-1-3: Clinician Survey

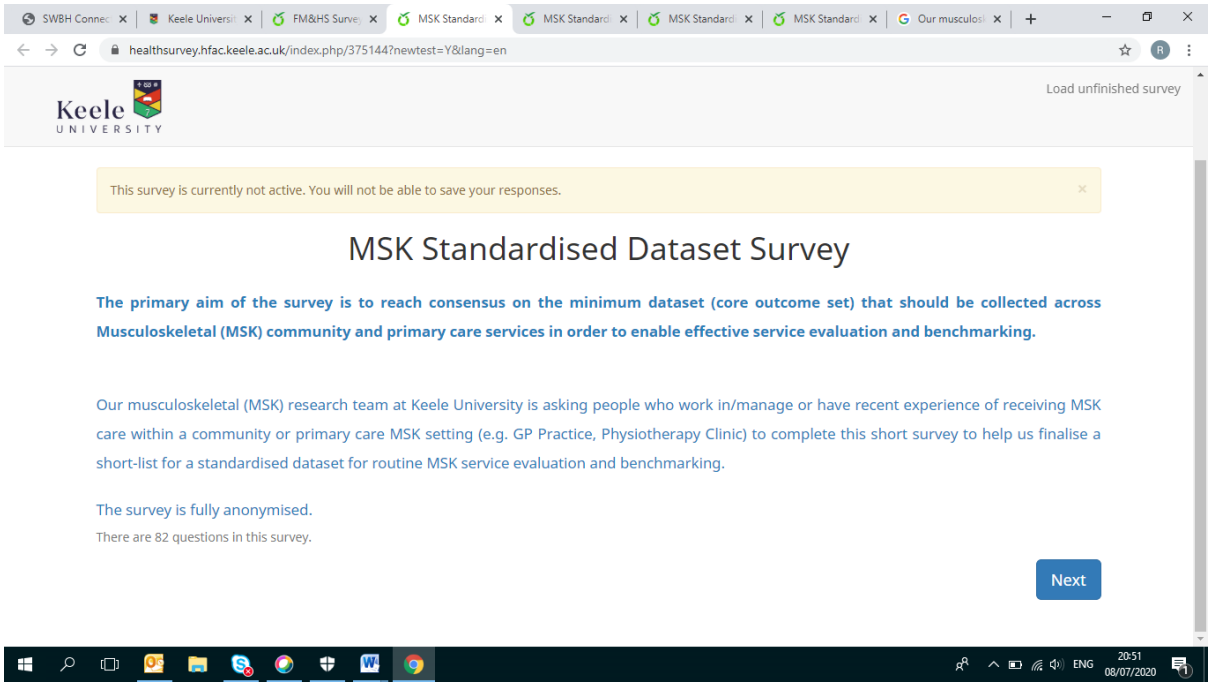
Clinician Survey

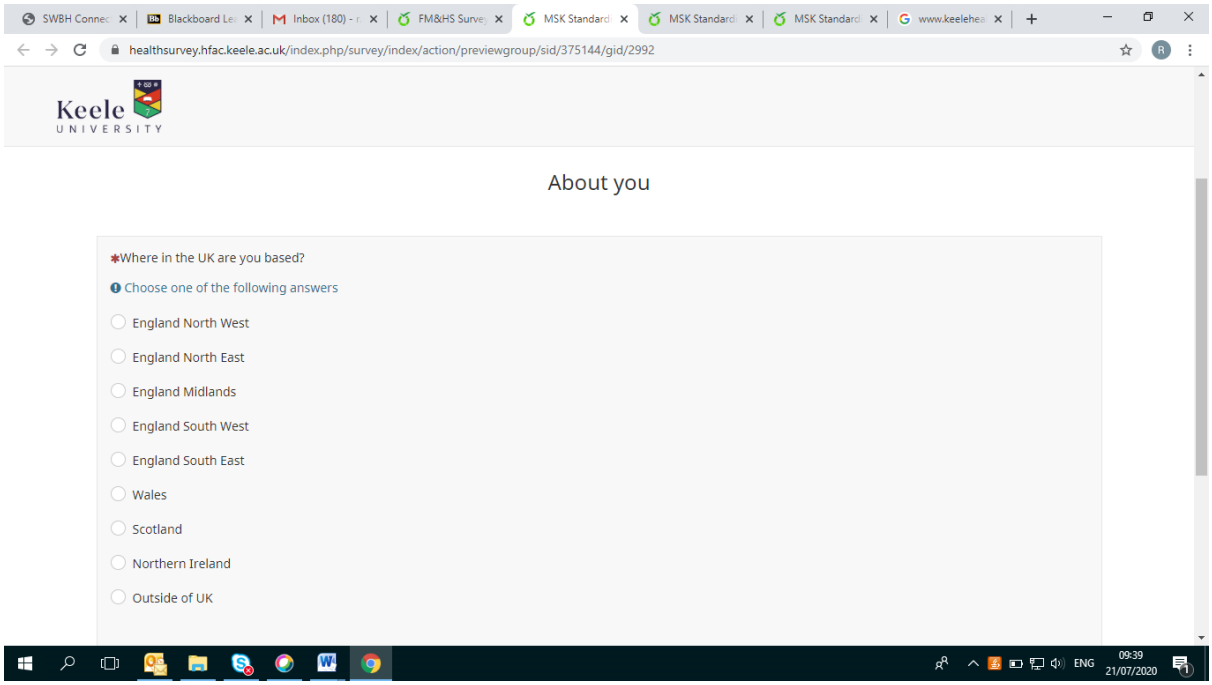
Introduction:

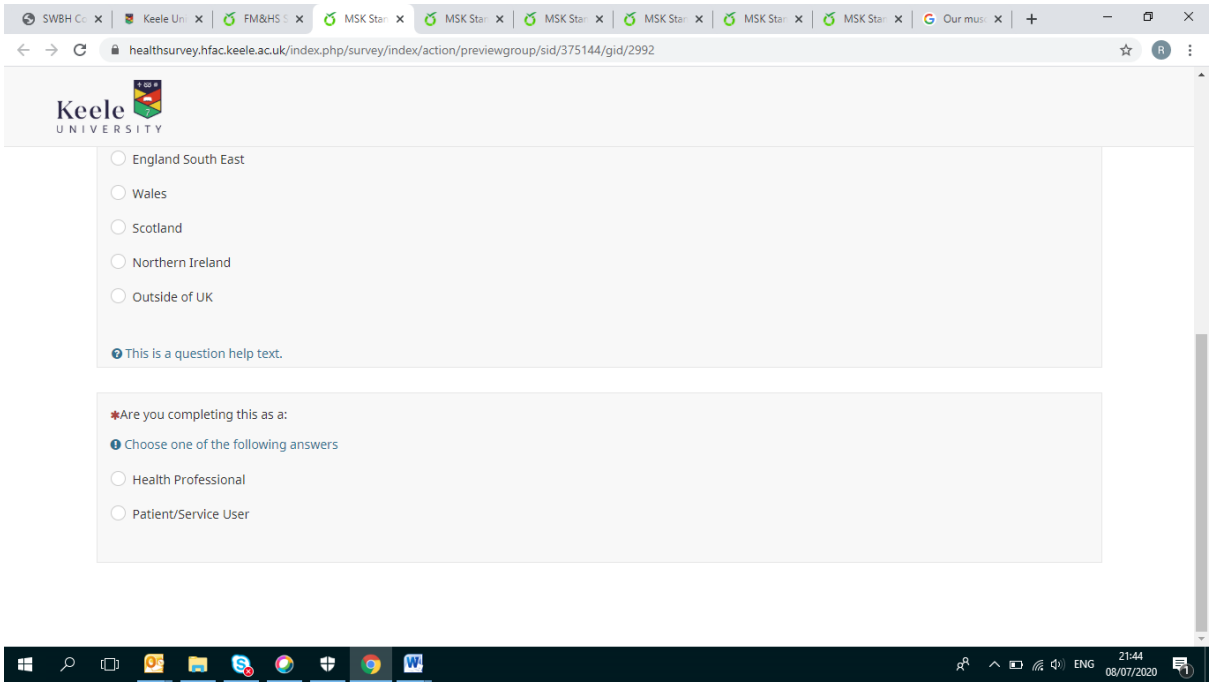
Short video: Dr Jonathan Hill introduction talking about the importance of; standardised data collection, the Musculoskeletal Health Questionnaire (MSK HQ), and the need for a national MSK audit.

Survey

Welcome All (Clinicians/Managers and Patients):





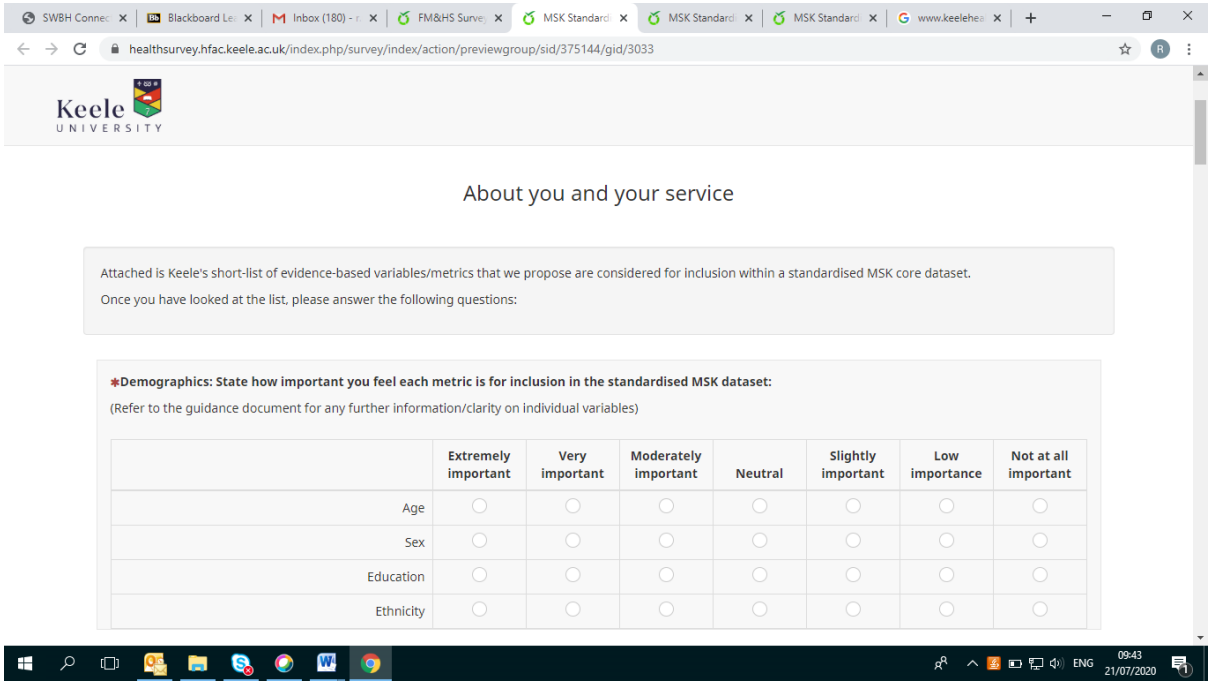


If the respondent ticks completing as a 'health professional' then it takes you to the following clinician survey:

Further Introduction: Roanna Burgess introducing to the online survey and the supporting material, and giving an overview of the project including brief aims, objectives and purpose of the survey and project.

Online material (proposed standardised dataset word document and excel coding document (these will be viewable from the Health Survey webpage and are currently available at: https://www.keele.ac.uk/pcsc/research/researchthemes/musculoskeletalpainandstratifiedcare/msktracker/ (bottom of webpage)

Clinician Survey



healthsurvey.hfac.keele.ac.uk/index.php/survey/index/action/previewgroup/sid/375144/gid/3033

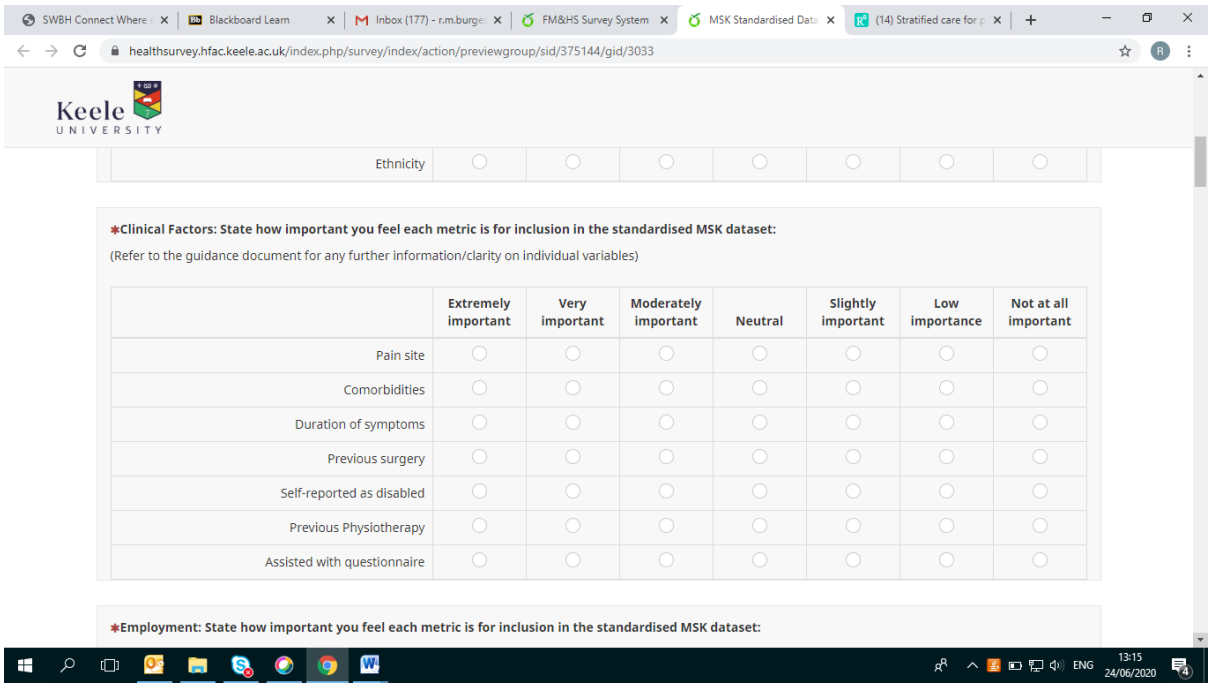
Keele UNIVERSITY

About you and your service

Attached is Keele's short-list of evidence-based variables/metrics that we propose are considered for inclusion within a standardised MSK core dataset.
Once you have looked at the list, please answer the following questions:

*Demographics: State how important you feel each metric is for inclusion in the standardised MSK dataset:
(Refer to the guidance document for any further information/clarity on individual variables)

	Extremely important	Very important	Moderately important	Neutral	Slightly important	Low importance	Not at all important
Age							
Sex							
Education							
Ethnicity							



healthsurvey.hfac.keele.ac.uk/index.php/survey/index/action/previewgroup/sid/375144/gid/3033

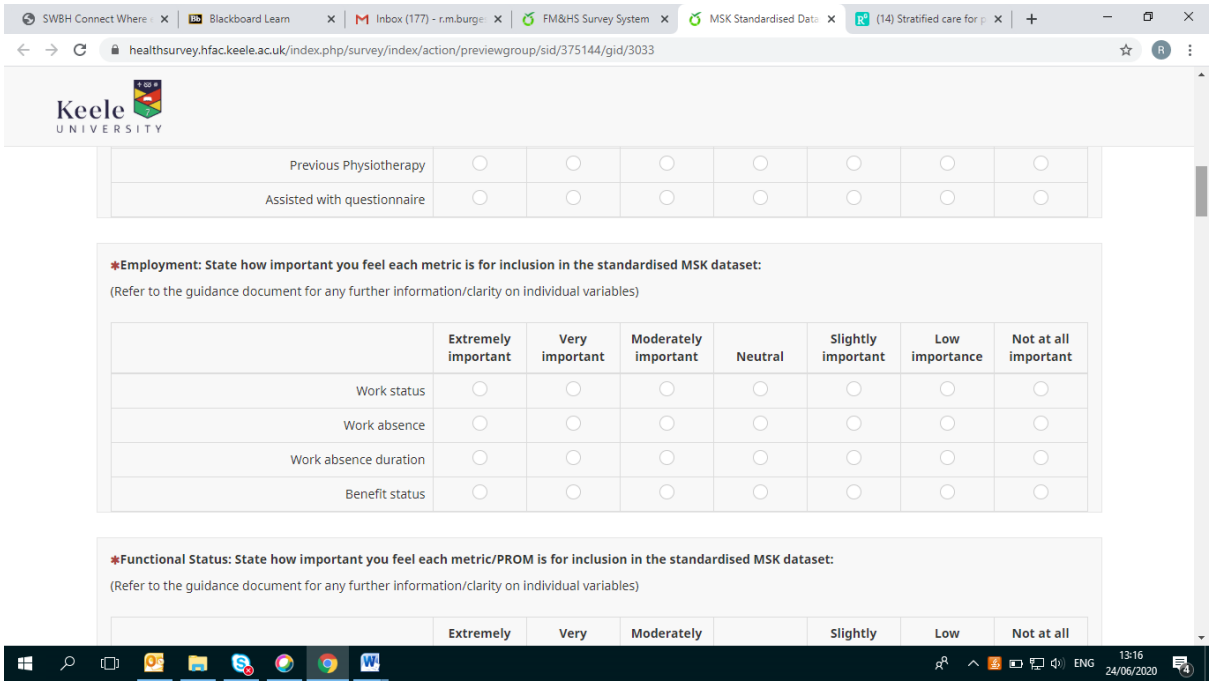
Keele UNIVERSITY

Ethnicity							

*Clinical Factors: State how important you feel each metric is for inclusion in the standardised MSK dataset:
(Refer to the guidance document for any further information/clarity on individual variables)

	Extremely important	Very important	Moderately important	Neutral	Slightly important	Low importance	Not at all important
Pain site							
Comorbidities							
Duration of symptoms							
Previous surgery							
Self-reported as disabled							
Previous Physiotherapy							
Assisted with questionnaire							

*Employment: State how important you feel each metric is for inclusion in the standardised MSK dataset:



Previous Physiotherapy	○	○	○	○	○	○	○
Assisted with questionnaire	○	○	○	○	○	○	○

***Employment: State how important you feel each metric is for inclusion in the standardised MSK dataset:**

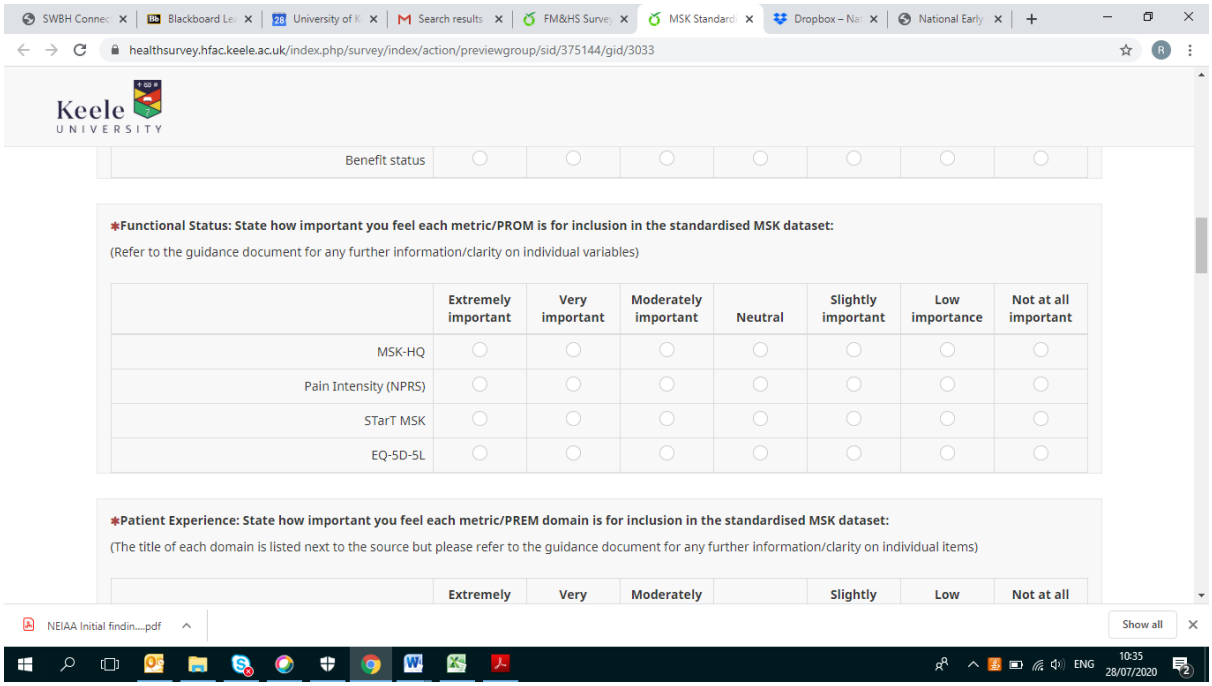
(Refer to the guidance document for any further information/clarity on individual variables)

	Extremely important	Very important	Moderately important	Neutral	Slightly important	Low importance	Not at all important
Work status	○	○	○	○	○	○	○
Work absence	○	○	○	○	○	○	○
Work absence duration	○	○	○	○	○	○	○
Benefit status	○	○	○	○	○	○	○

***Functional Status: State how important you feel each metric/PROM is for inclusion in the standardised MSK dataset:**

(Refer to the guidance document for any further information/clarity on individual variables)

	Extremely	Very	Moderately		Slightly	Low	Not at all



Benefit status	○	○	○	○	○	○	○

***Functional Status: State how important you feel each metric/PROM is for inclusion in the standardised MSK dataset:**

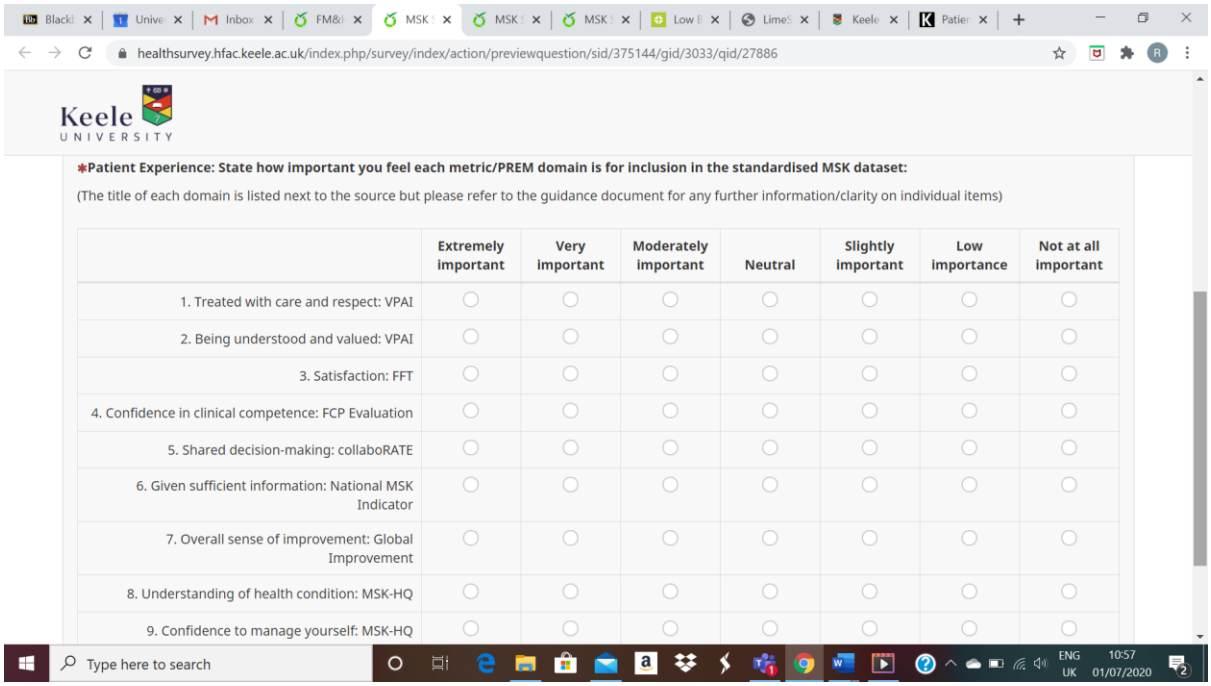
(Refer to the guidance document for any further information/clarity on individual variables)

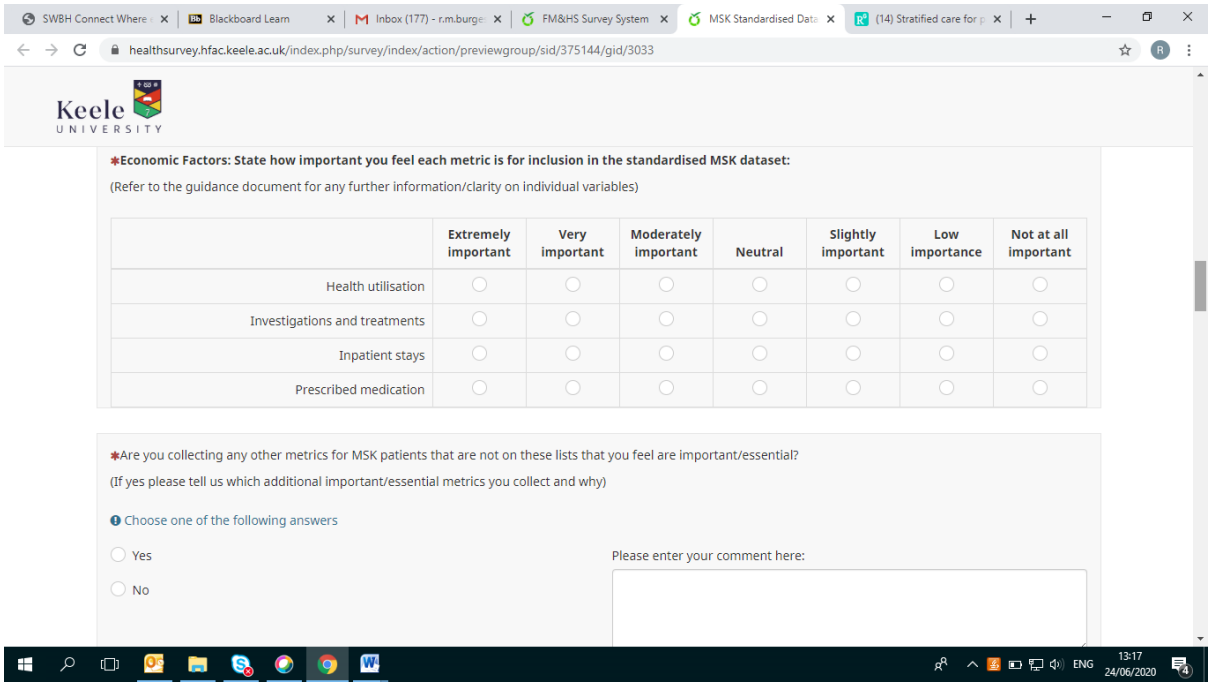
	Extremely important	Very important	Moderately important	Neutral	Slightly important	Low importance	Not at all important
MSK-HQ	○	○	○	○	○	○	○
Pain Intensity (NPRS)	○	○	○	○	○	○	○
STarT MSK	○	○	○	○	○	○	○
EQ-5D-5L	○	○	○	○	○	○	○

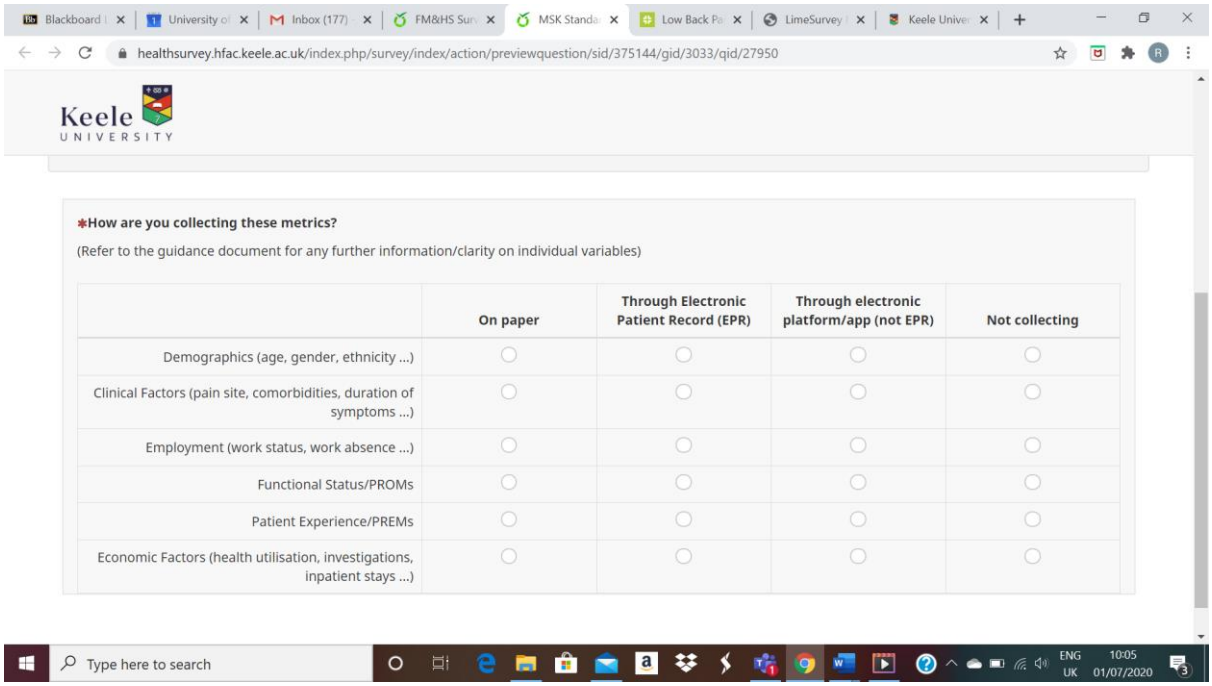
***Patient Experience: State how important you feel each metric/PREM domain is for inclusion in the standardised MSK dataset:**

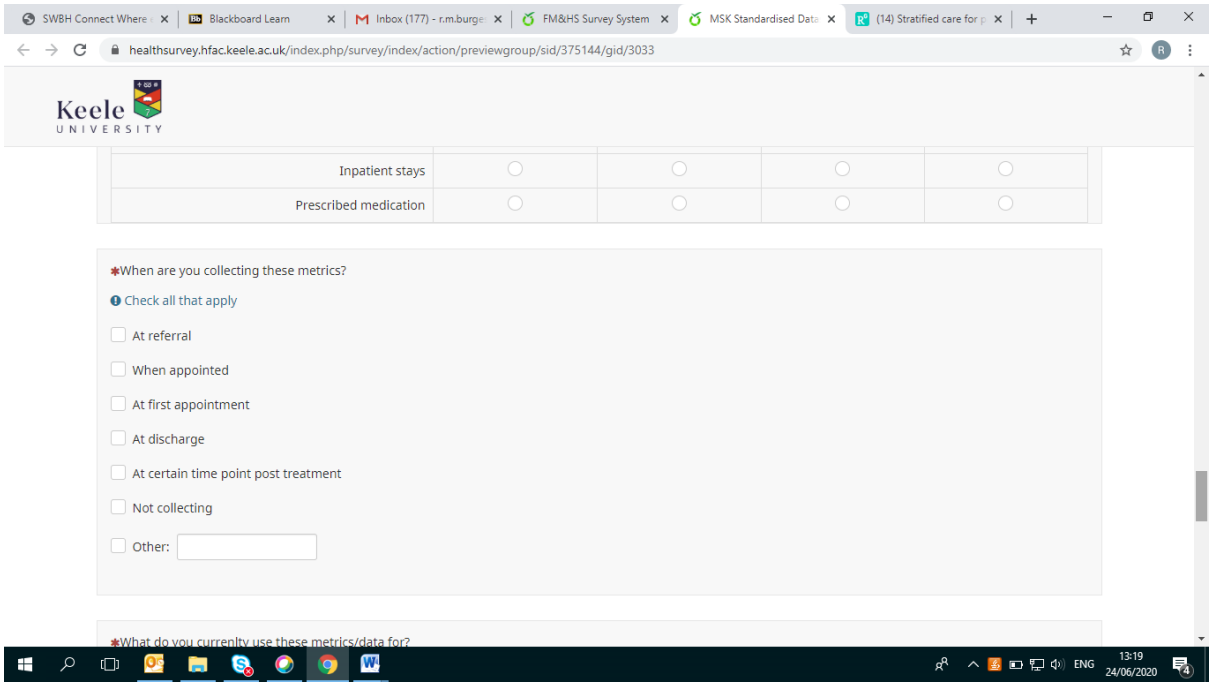
(The title of each domain is listed next to the source but please refer to the guidance document for any further information/clarity on individual items)

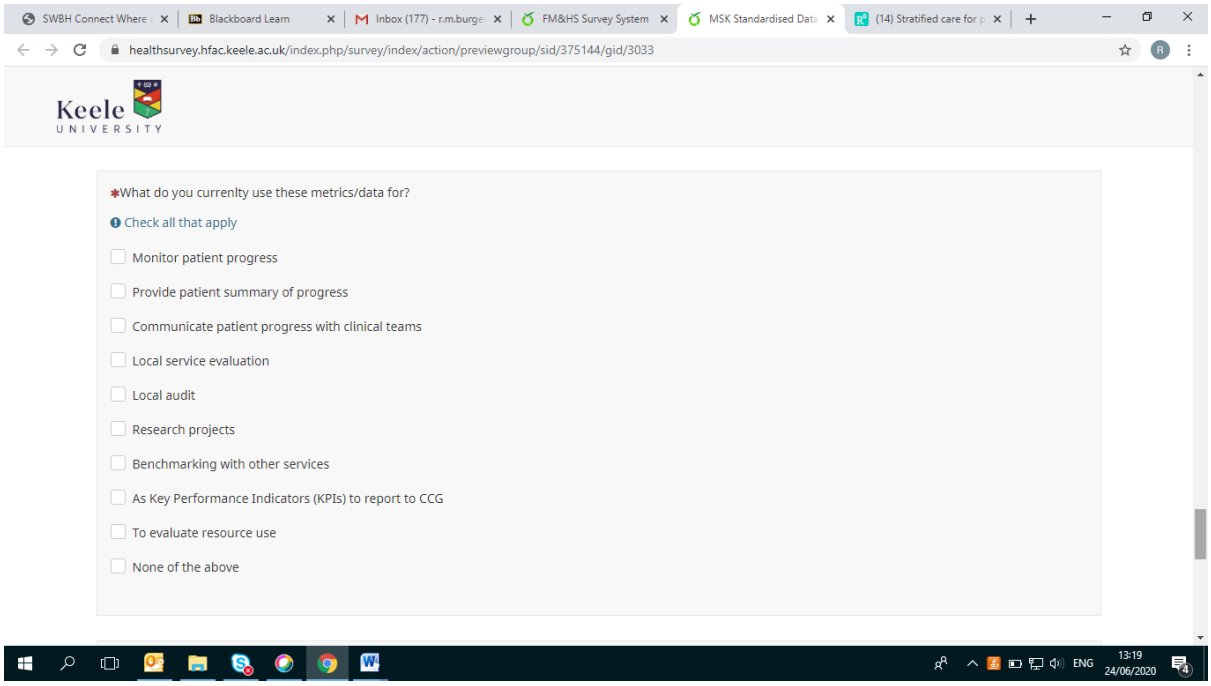
	Extremely	Very	Moderately		Slightly	Low	Not at all

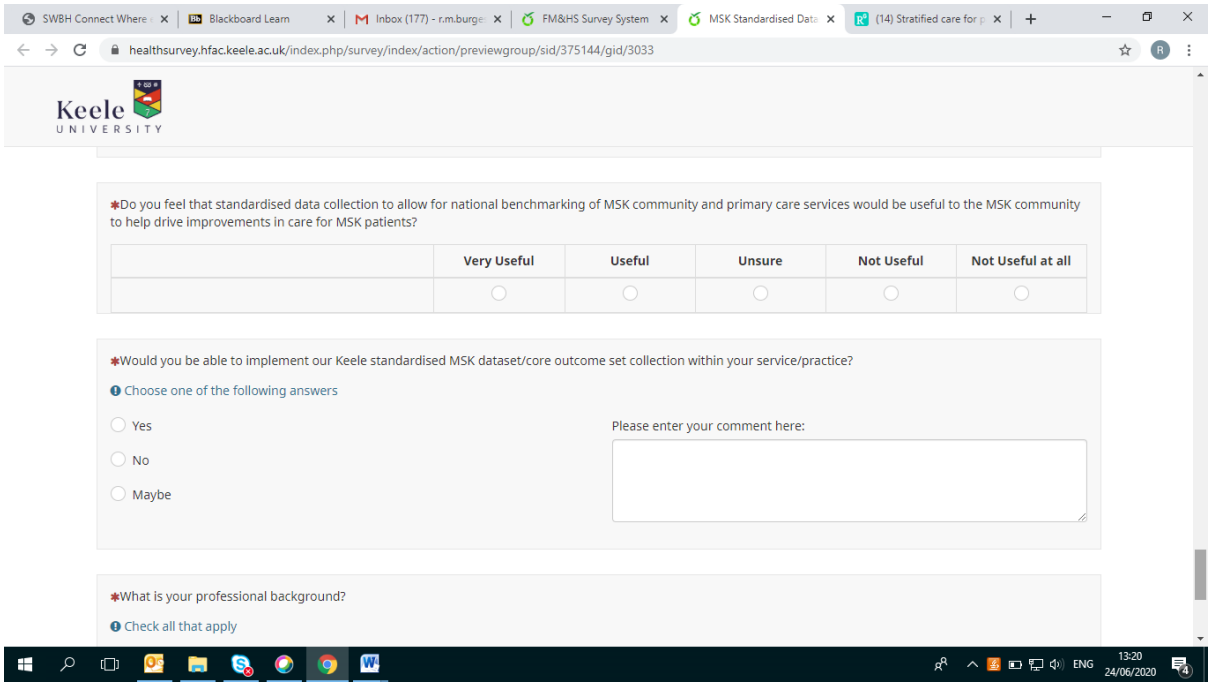

healthsurvey.hfac.keele.ac.uk/index.php/survey/index/action/previewquestion/sid/375144/gid/3033/qid/27886
Keele UNIVERSITY
*Patient Experience: State how important you feel each metric/PREM domain is for inclusion in the standardised MSK dataset:
(The title of each domain is listed next to the source but please refer to the guidance document for any further information/clarity on individual items)
Extremely important
Very important
Moderately important
Neutral
Slightly important
Low importance
Not at all important
1. Treated with care and respect: VPAI
2. Being understood and valued: VPAI
3. Satisfaction: FFT
4. Confidence in clinical competence: FCP Evaluation
5. Shared decision-making: collaboRATE
6. Given sufficient information: National MSK Indicator
7. Overall sense of improvement: Global Improvement
8. Understanding of health condition: MSK-HQ
9. Confidence to manage yourself: MSK-HQ
Type here to search
ENG UK 10:57 01/07/2020

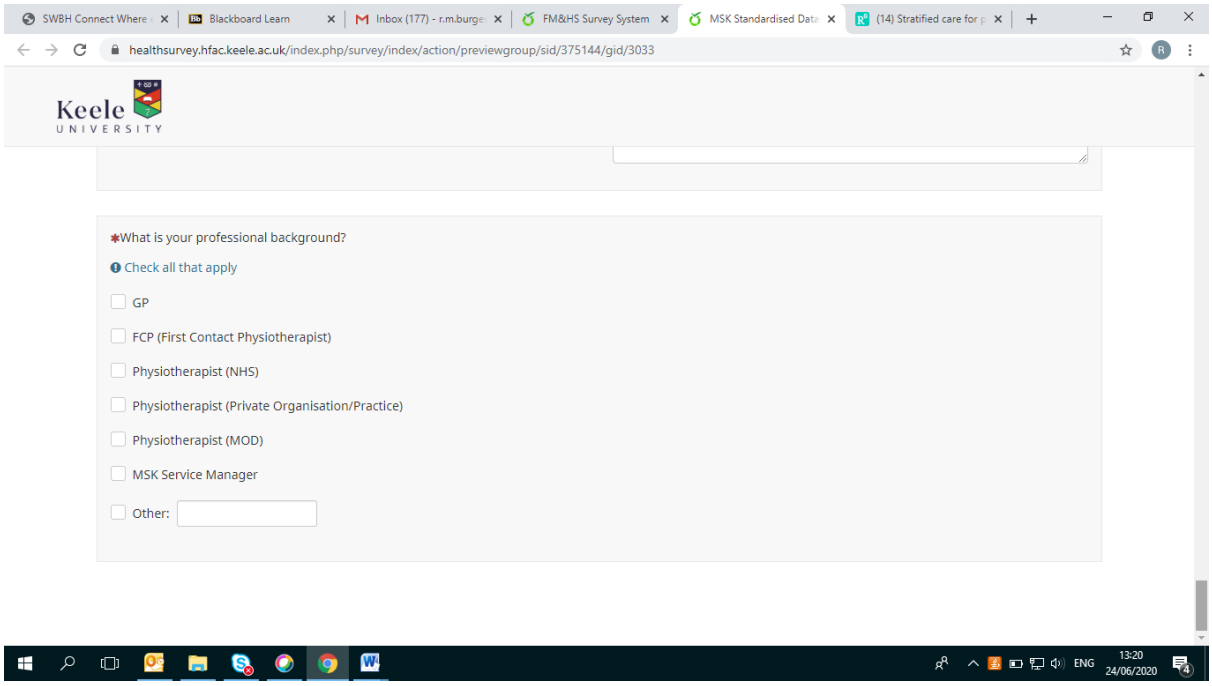

healthsurvey.hfac.keele.ac.uk/index.php/survey/index/action/previewgroup/sid/375144/gid/3033
Keele UNIVERSITY
*Economic Factors: State how important you feel each metric is for inclusion in the standardised MSK dataset:
(Refer to the guidance document for any further information/clarity on individual variables)
Extremely important
Very important
Moderately important
Neutral
Slightly important
Low importance
Not at all important
Health utilisation
Investigations and treatments
Inpatient stays
Prescribed medication
*Are you collecting any other metrics for MSK patients that are not on these lists that you feel are important/essential?
(If yes please tell us which additional important/essential metrics you collect and why)
Choose one of the following answers
Yes
No
Please enter your comment here:
ENG 13:17 24/06/2020


healthsurvey.hfac.keele.ac.uk/index.php/survey/index/action/previewquestion/sid/375144/gid/3033/qid/27950
Keele UNIVERSITY
*How are you collecting these metrics?
(Refer to the guidance document for any further information/clarity on individual variables)
On paper
Through Electronic Patient Record (EPR)
Through electronic platform/app (not EPR)
Not collecting
Demographics (age, gender, ethnicity ...)
Clinical Factors (pain site, comorbidities, duration of symptoms ...)
Employment (work status, work absence ...)
Functional Status/PROMs
Patient Experience/PREMs
Economic Factors (health utilisation, investigations, inpatient stays ...)
Type here to search
ENG UK 10:05 01/07/2020


healthsurvey.hfac.keele.ac.uk/index.php/survey/index/action/previewgroup/sid/375144/gid/3033
Keele UNIVERSITY
Inpatient stays
Prescribed medication
*When are you collecting these metrics?
Check all that apply
At referral
When appointed
At first appointment
At discharge
At certain time point post treatment
Not collecting
Other:
*What do you currently use these metrics/data for?
ENG 13:19 24/06/2020


SWBH Connect Where
Blackboard Learn
Inbox (177) - r.m.burge
FM&HS Survey System
MSK Standardised Data
(14) Stratified care for
healthsurvey.hfac.keele.ac.uk/index.php/survey/index/action/previewgroup/sid/375144/gid/3033
Keele
UNIVERSITY
*What do you currenlty use these metrics/data for?
Check all that apply
Monitor patient progress
Provide patient summary of progress
Communicate patient progress with clinical teams
Local service evaluation
Local audit
Research projects
Benchmarking with other services
As Key Performance Indicators (KPIs) to report to CCG
To evaluate resource use
None of the above
ENG
13:19
24/06/2020


SWBH Connect Where
Blackboard Learn
Inbox (177) - r.m.burge
FM&HS Survey System
MSK Standardised Data
(14) Stratified care for
healthsurvey.hfac.keele.ac.uk/index.php/survey/index/action/previewgroup/sid/375144/gid/3033
Keele
UNIVERSITY
*Do you feel that standardised data collection to allow for national benchmarking of MSK community and primary care services would be useful to the MSK community to help drive improvements in care for MSK patients?
Very Useful
Useful
Unsure
Not Useful
Not Useful at all
*Would you be able to implement our Keele standardised MSK dataset/core outcome set collection within your service/practice?
Choose one of the following answers
Yes
No
Maybe
Please enter your comment here:
*What is your professional background?
Check all that apply
ENG
13:20
24/06/2020



END MESSAGE

Thank you for your time. A summary of results and updates to the MSK Standardised Dataset and associated documents will be available as they are released on Keele University's website:

https://www.keele.ac.uk/pcsc/research/researchthemes/musculoskeletalpainandstratifiedcare/msktracker/

Appendix 5-1-4: Patient Survey

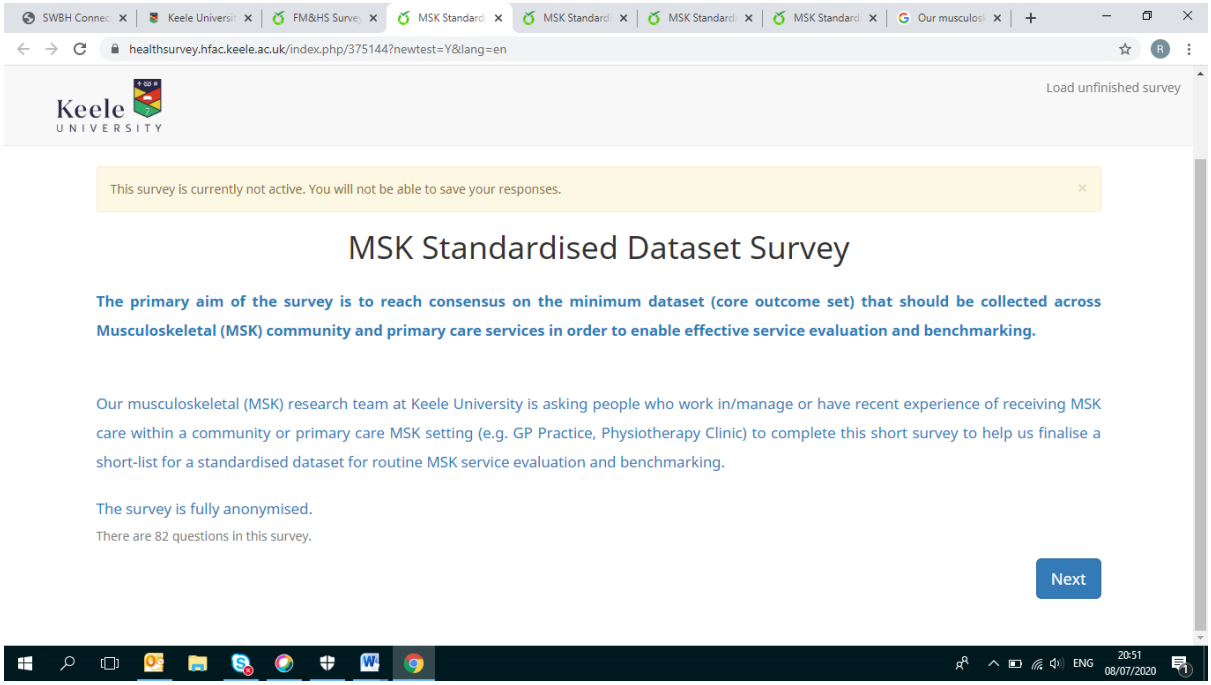
Patient Survey

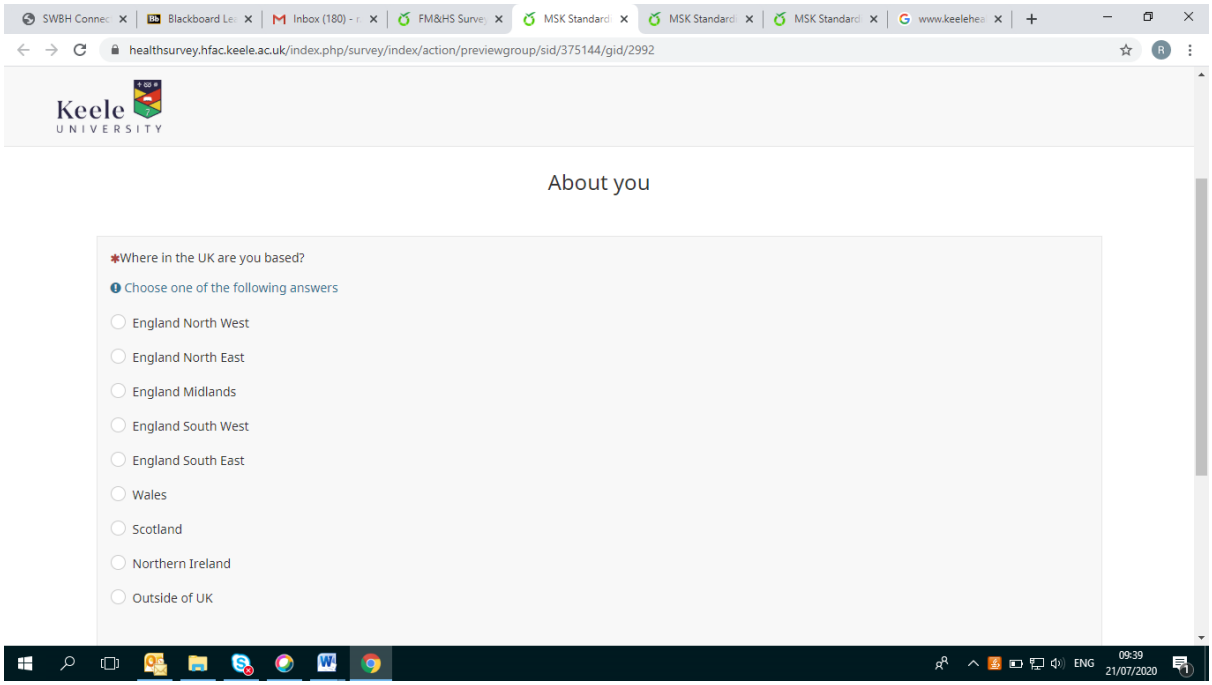
Introduction

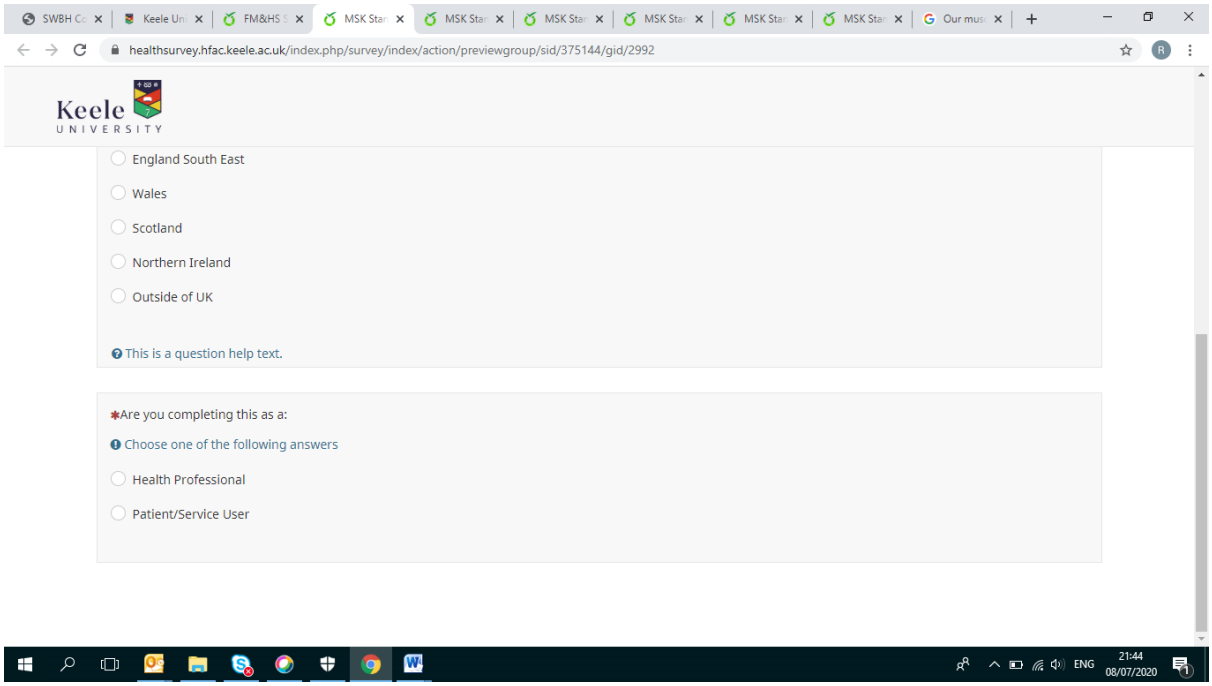
Short video: Dr Jonathan Hill introduction talking about the importance of; standardised data collection, the Musculoskeletal Health Questionnaire (MSK HQ), and the need for national MSK audit.

Survey

All (Clinicians/Managers and Patients)



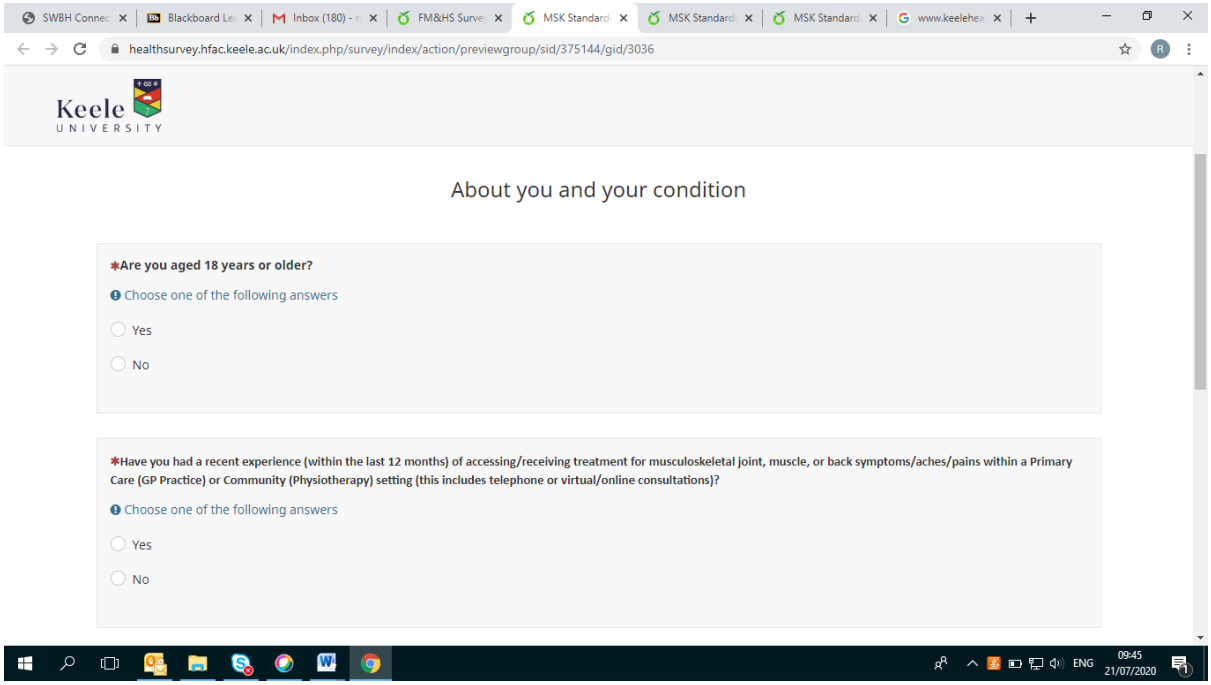




If the respondent ticks completing as a 'patient/service user' then it takes you to the following patient survey:

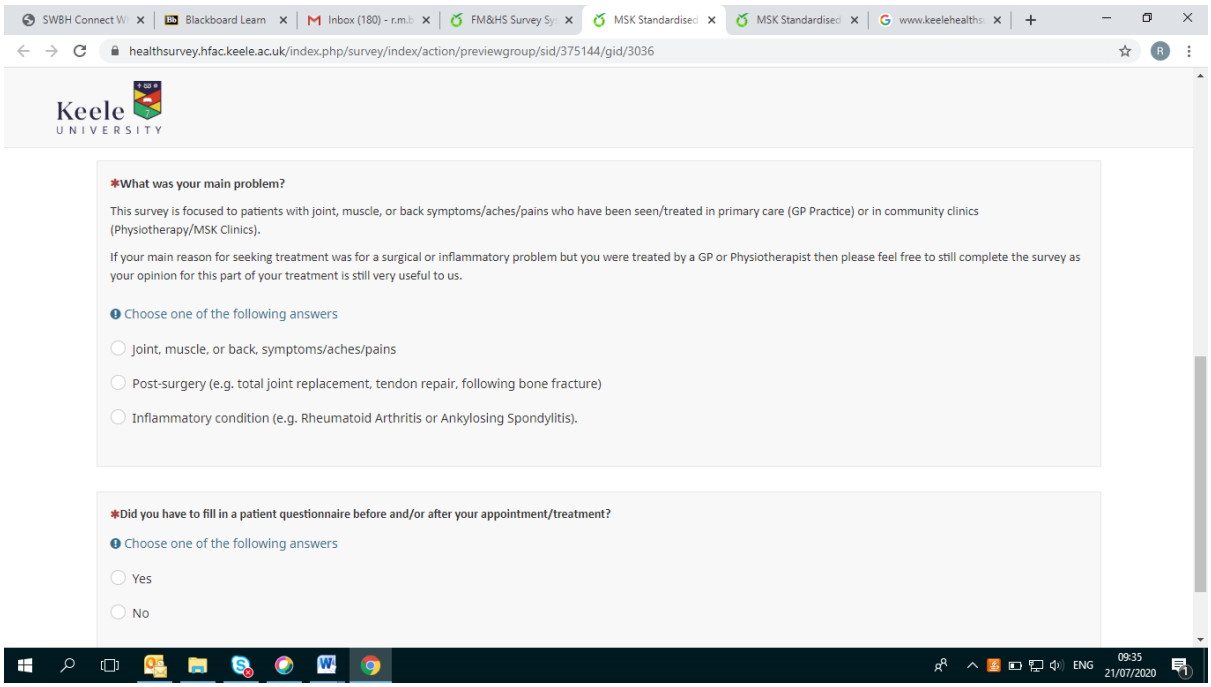
Further introduction: this will include a short video for patients (Roanna Burgess) outlining the aims of the project and survey, how data can improve patient care and how they can help. This will be followed by taking them through the format of the survey. Patients will see a demo of the proposed MSK patient questionnaire in sections and then answer questions on what they think, and then answer some general questions on the entire questionnaire, they will also be talked through inclusion criteria (need to be 18 or over and have had a recent experience (last 12 months) of visiting their GP or Physiotherapist for a muscle, joint, or back related problem), and reassured that their feedback is fully anonymised.

About you and your condition:



If the patient answers no to either of these questions they will exit the survey with a message:

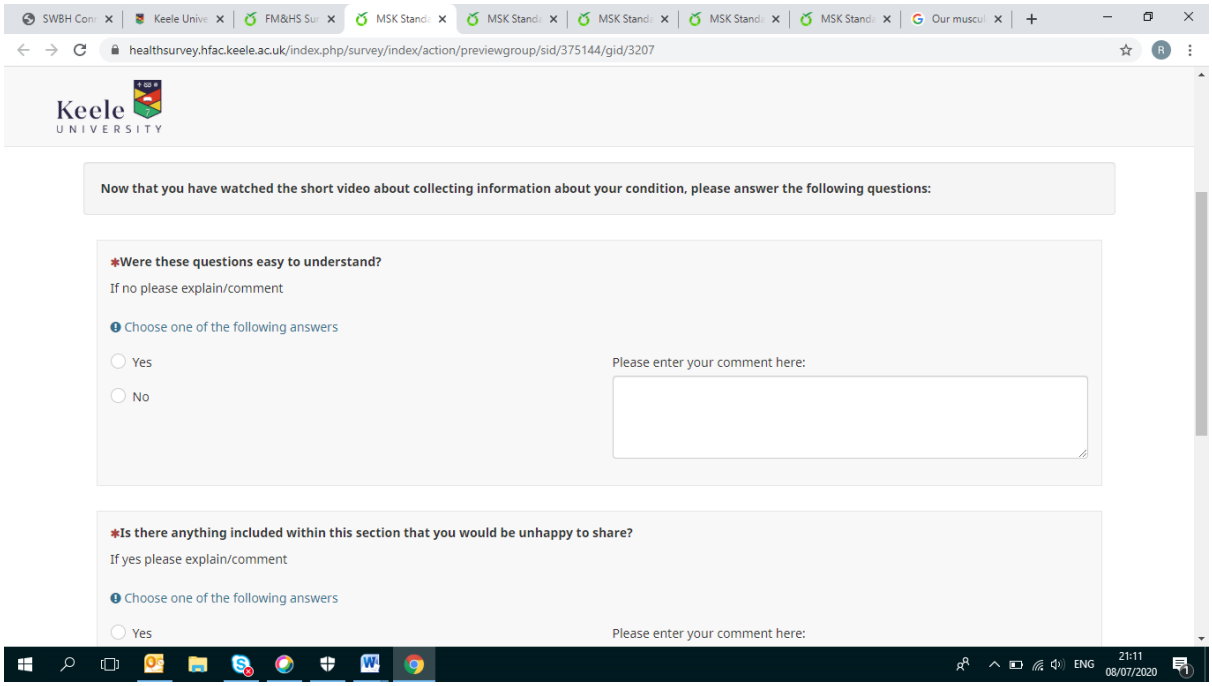
'Thank you for your interest in taking part in the Keele MSK Data Patient Survey, unfortunately you do not meet the criteria to be included this time but we are very grateful for your interest.'

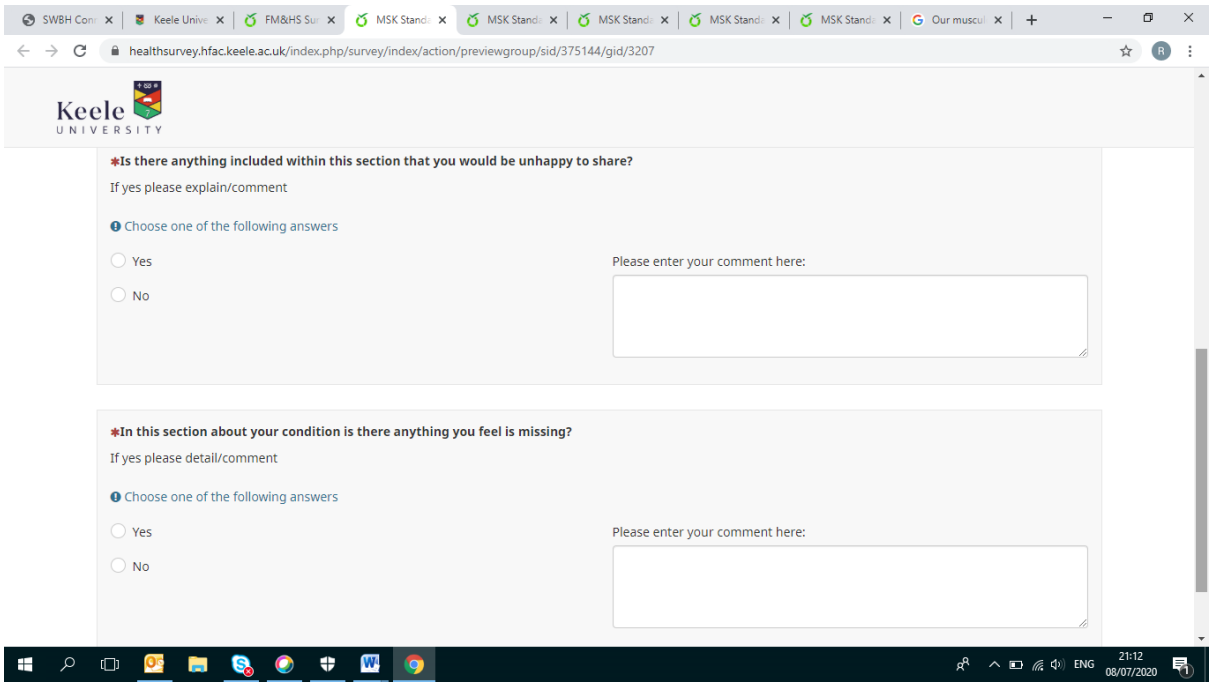


1. **Clinical Factors (About your condition)**

Patients will view a video clip demo of the proposed MSK questionnaire relating to their MSK condition (see Appendix 5 (Section 1) for detail).

Questions

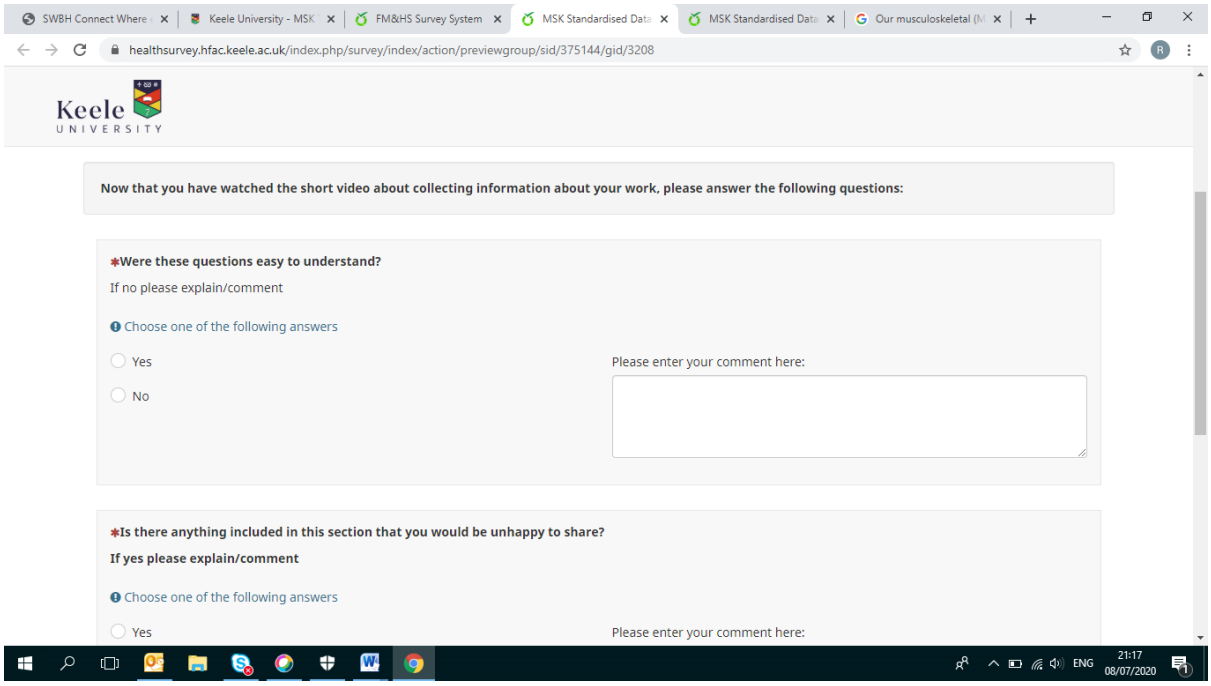


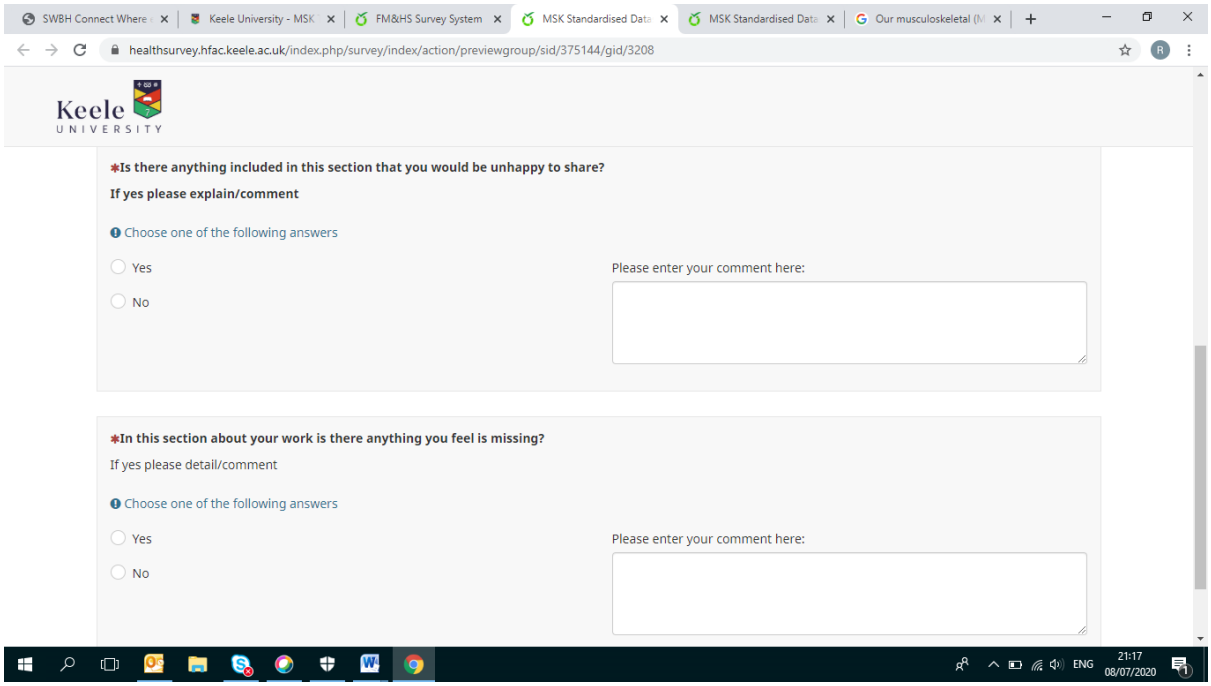


2. **About your work**

Patients will view a video clip demo of the proposed MSK questionnaire relating to their work (see Appendix 5 (Section 2) for detail).

Work Questions

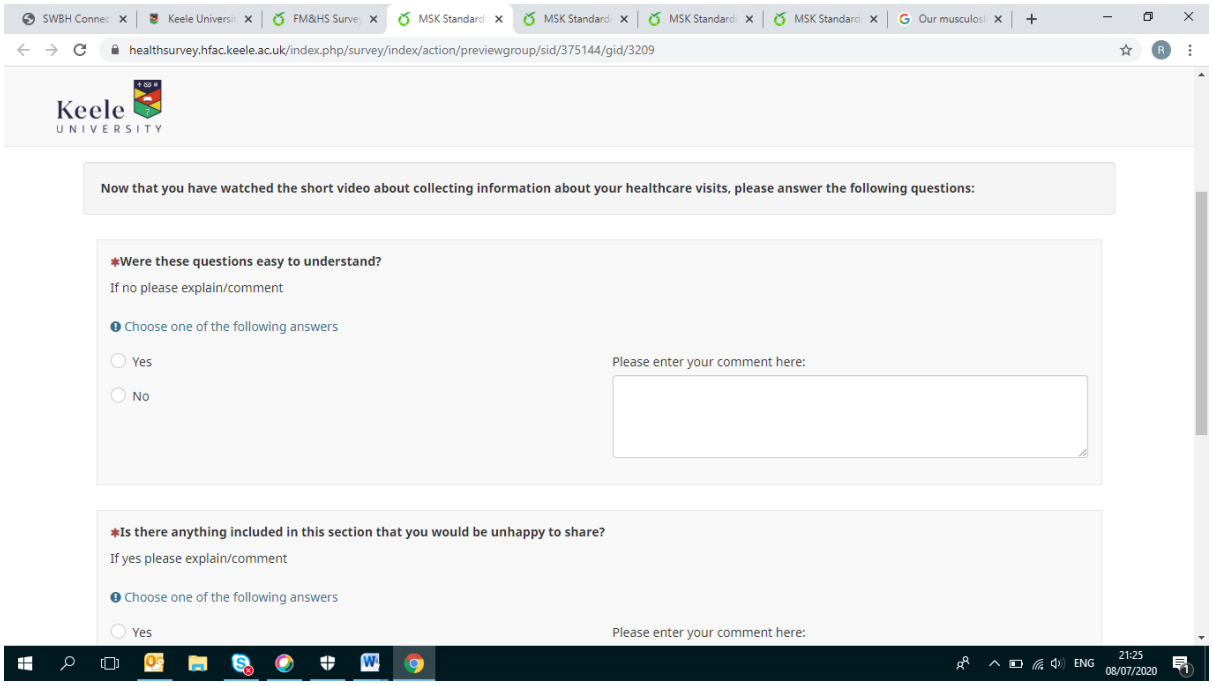


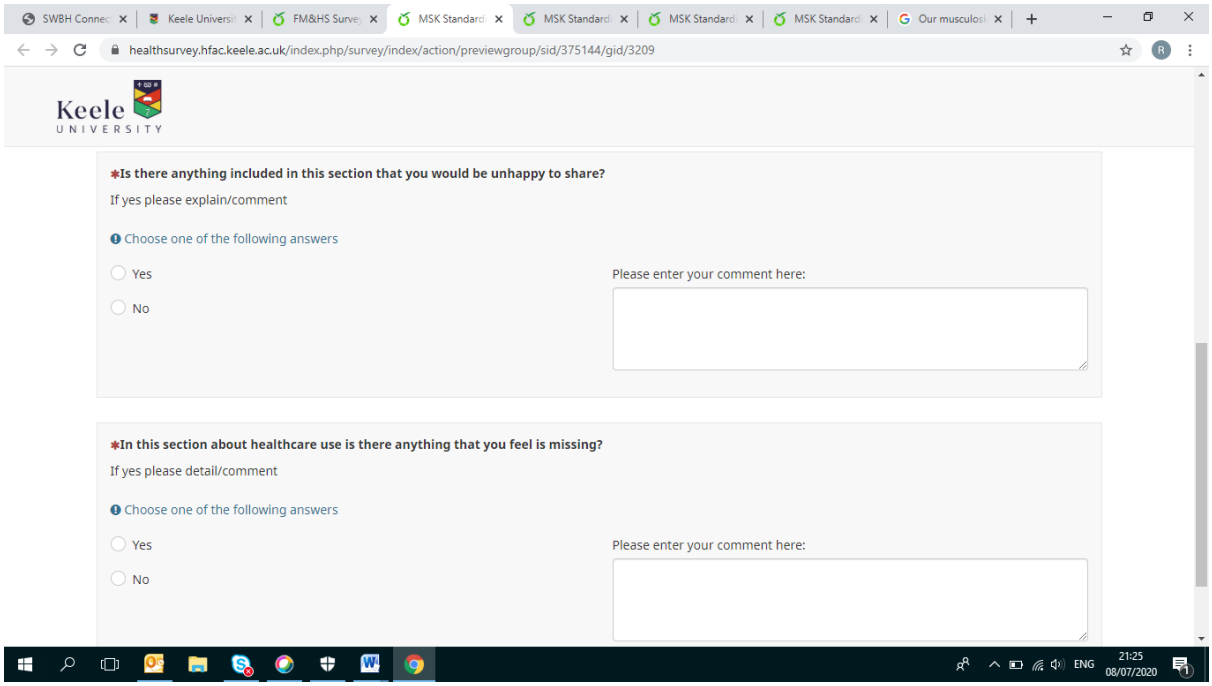


3. About your healthcare visits

Patients will view a video clip demo of the proposed MSK questionnaire relating to their healthcare visits (see Appendix 5 (Section 3) for detail).

Healthcare visits questions

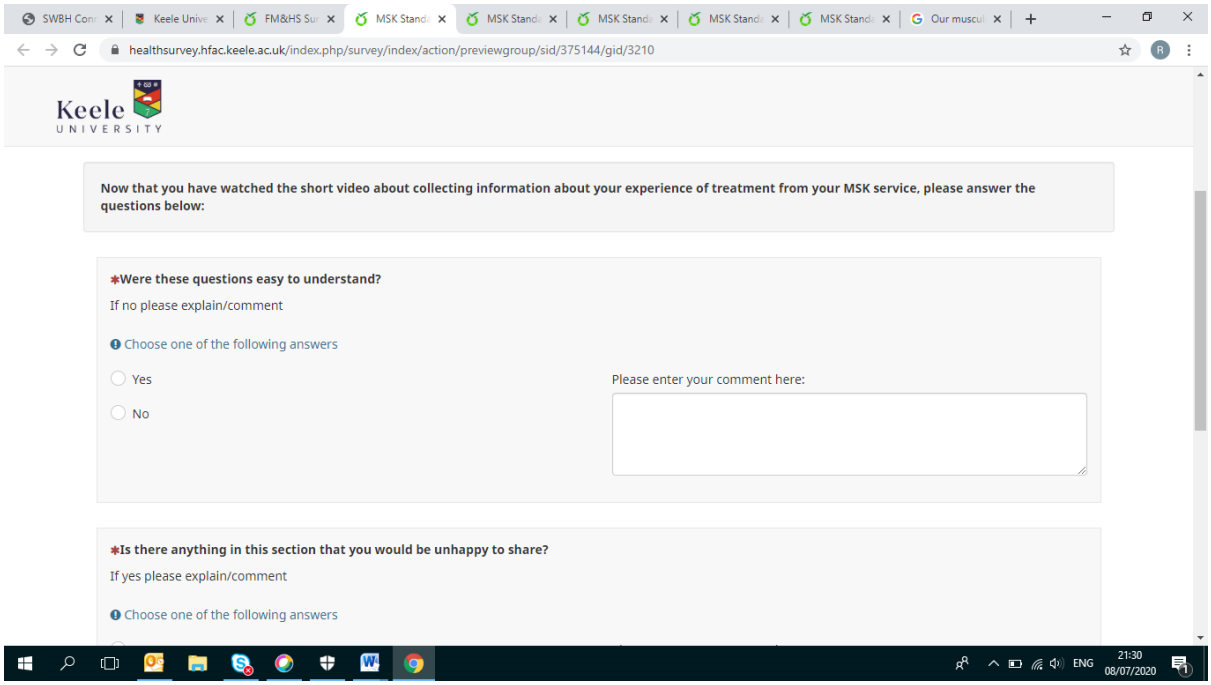


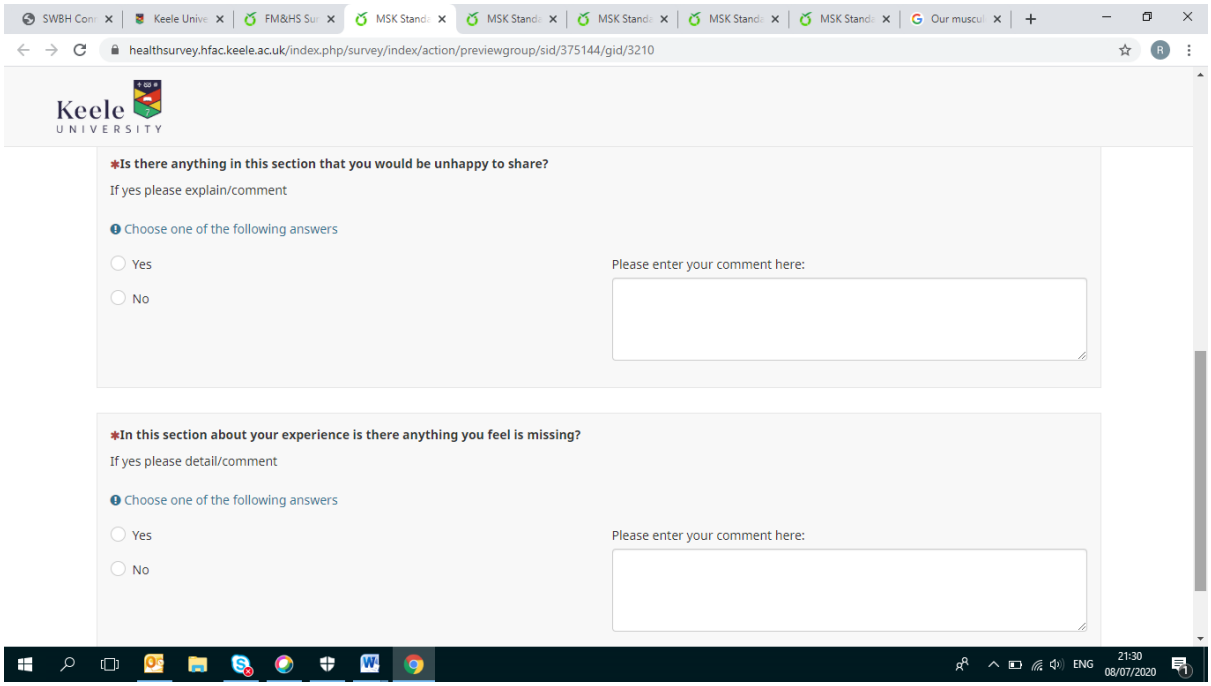


4. Patient Experience (About your experience)

Patients will view a video clip demo of the proposed MSK questionnaire relating to their experience of treatment within their MSK service (see Appendix 5 (Section 4) for detail).

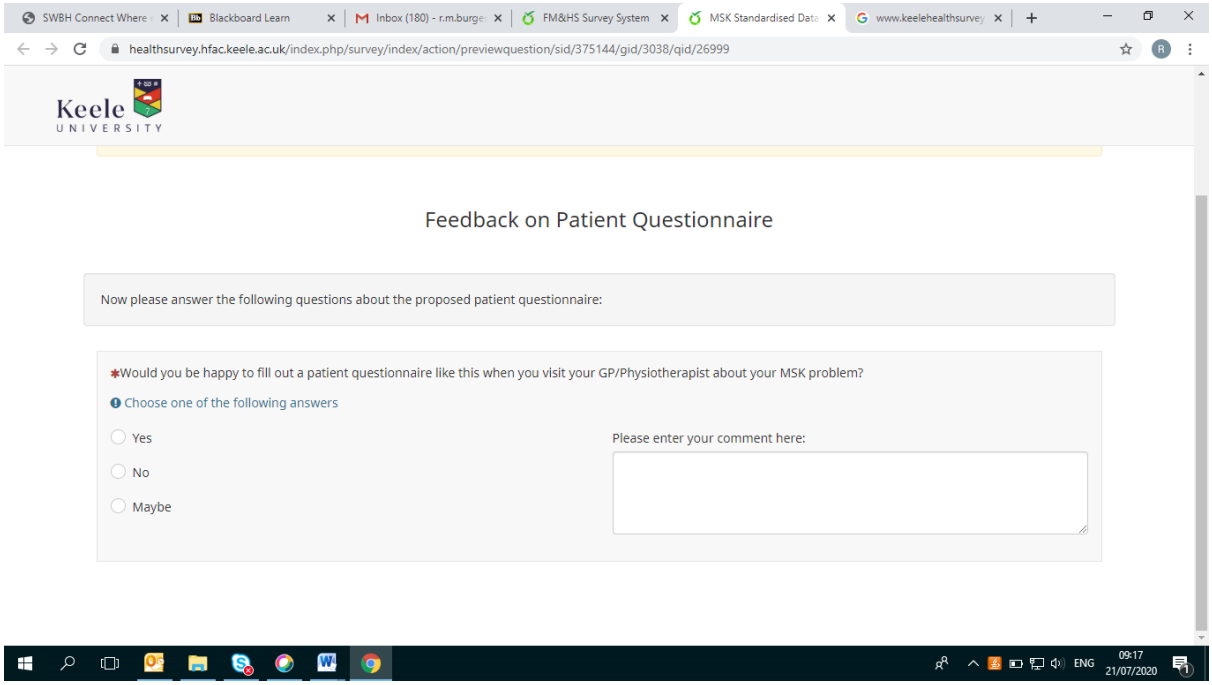
Patient Experience Questions:

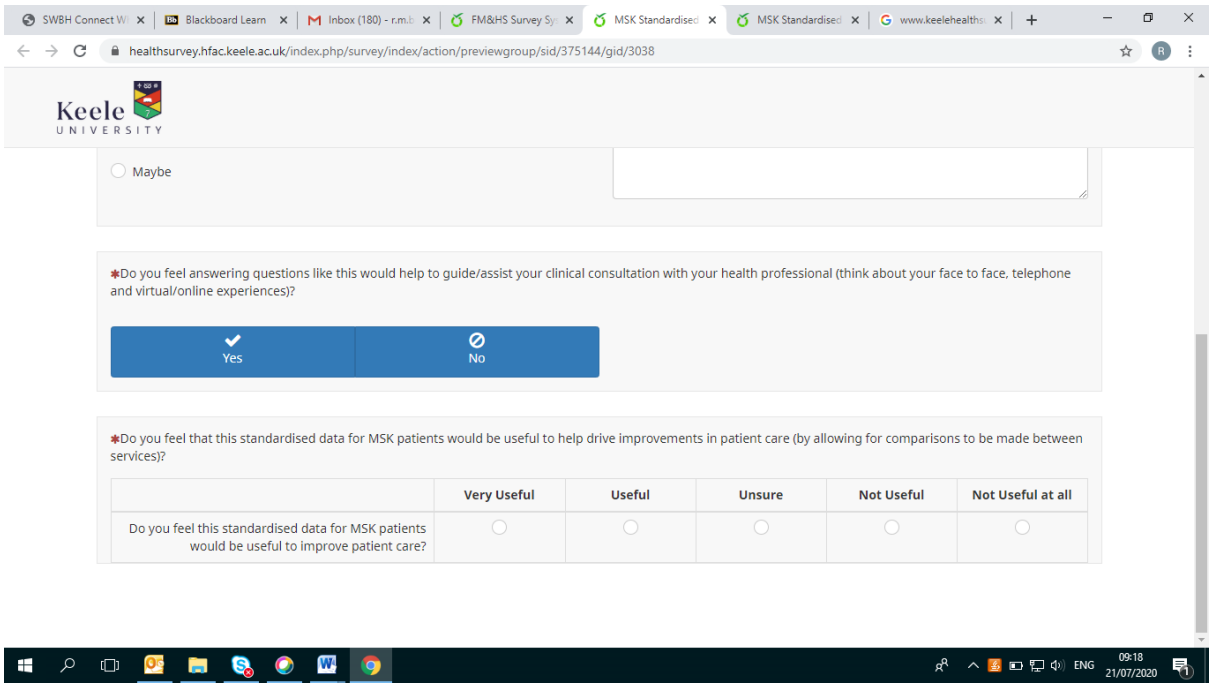




5. Final Questions

Feedback on proposed patient questionnaire:





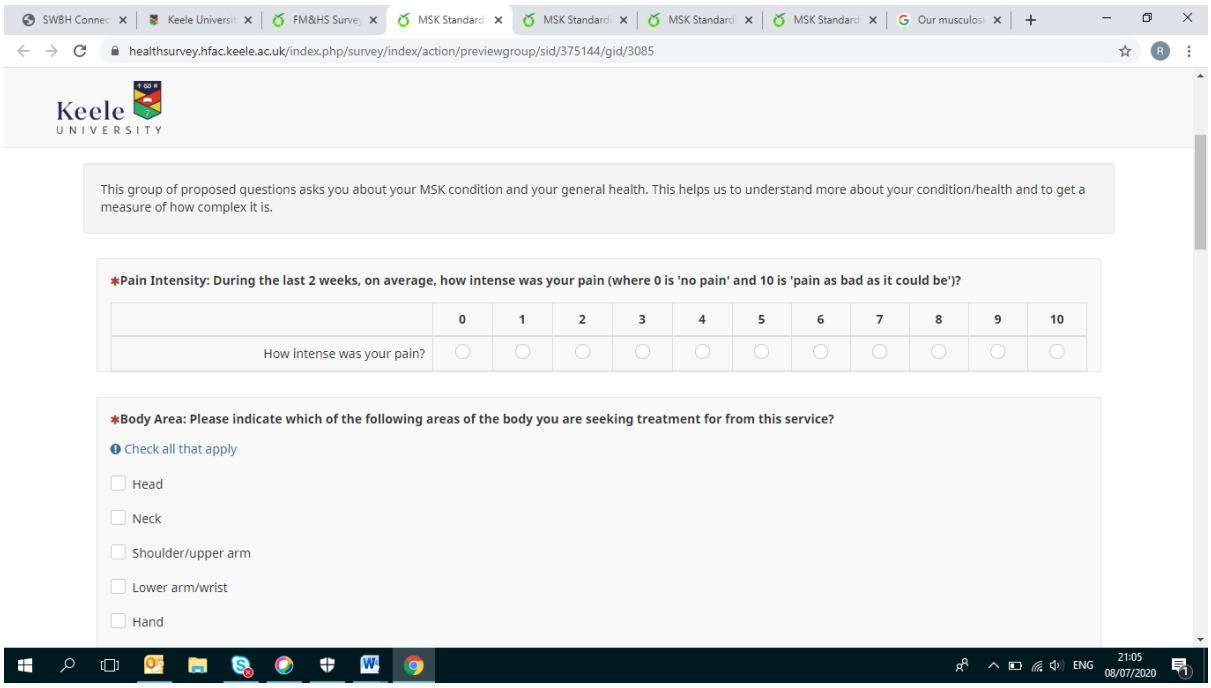
End of survey

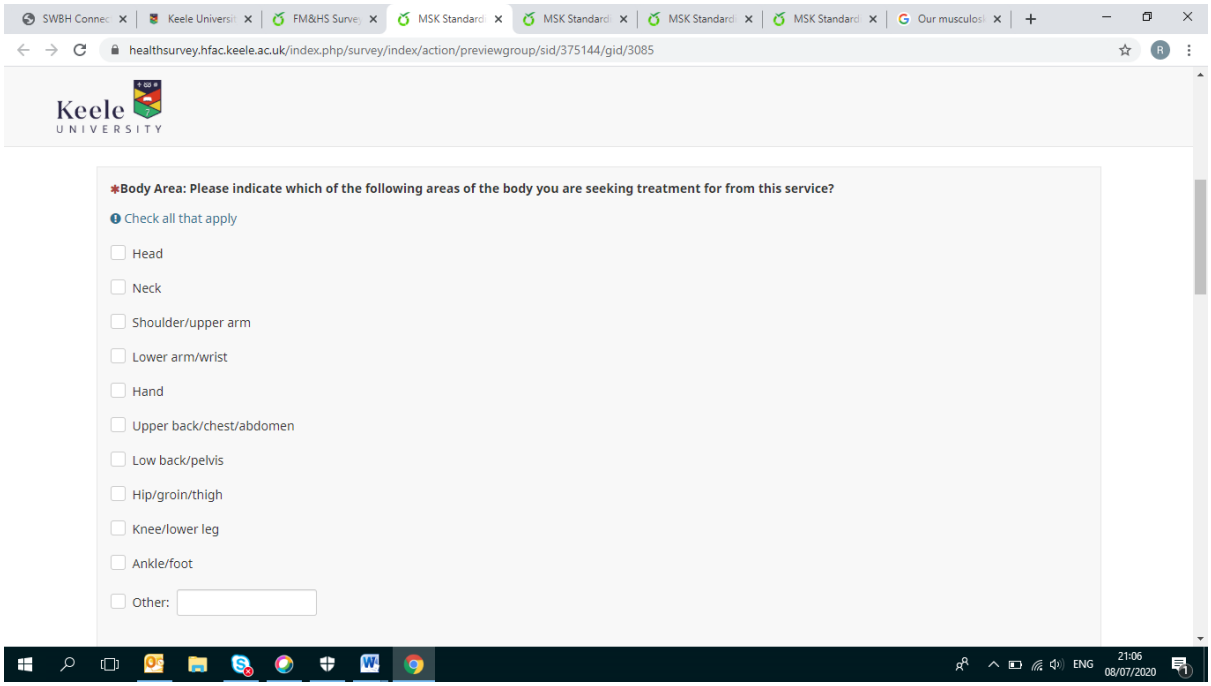
End message 'Thank you for completing the Keele MSK Data Patient Survey, we really value your opinion and your time in helping us improve NHS care for the future.'

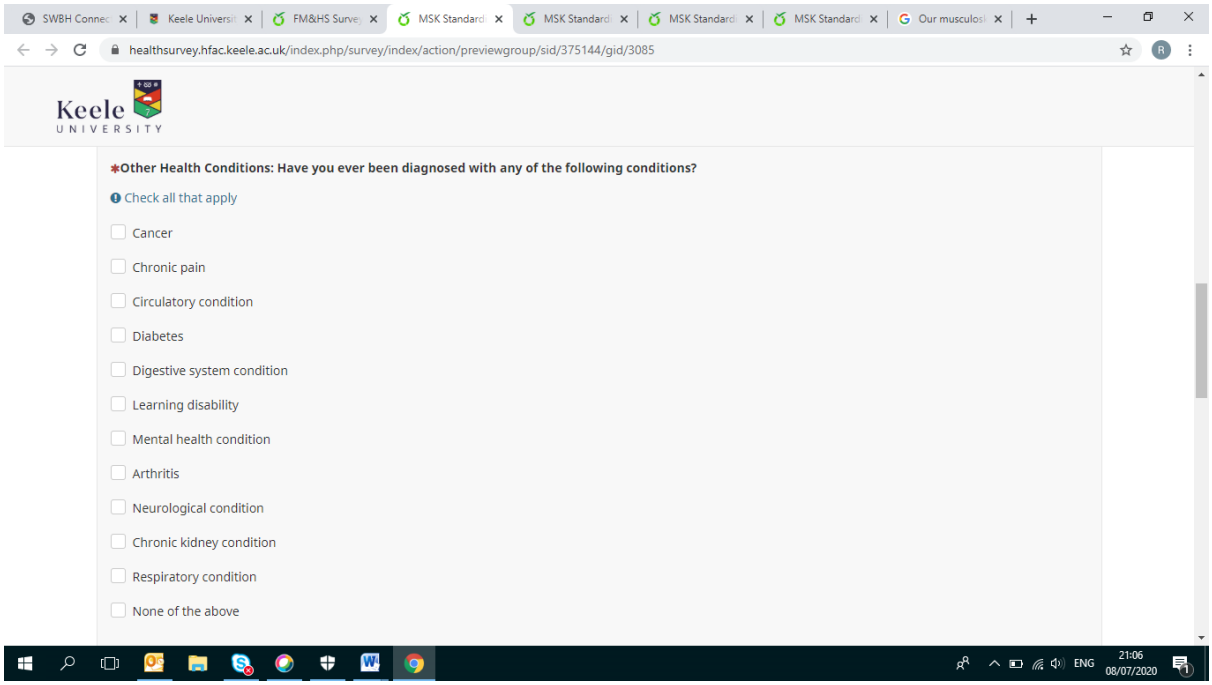
Appendix 5-1-5: Patient Survey Detail

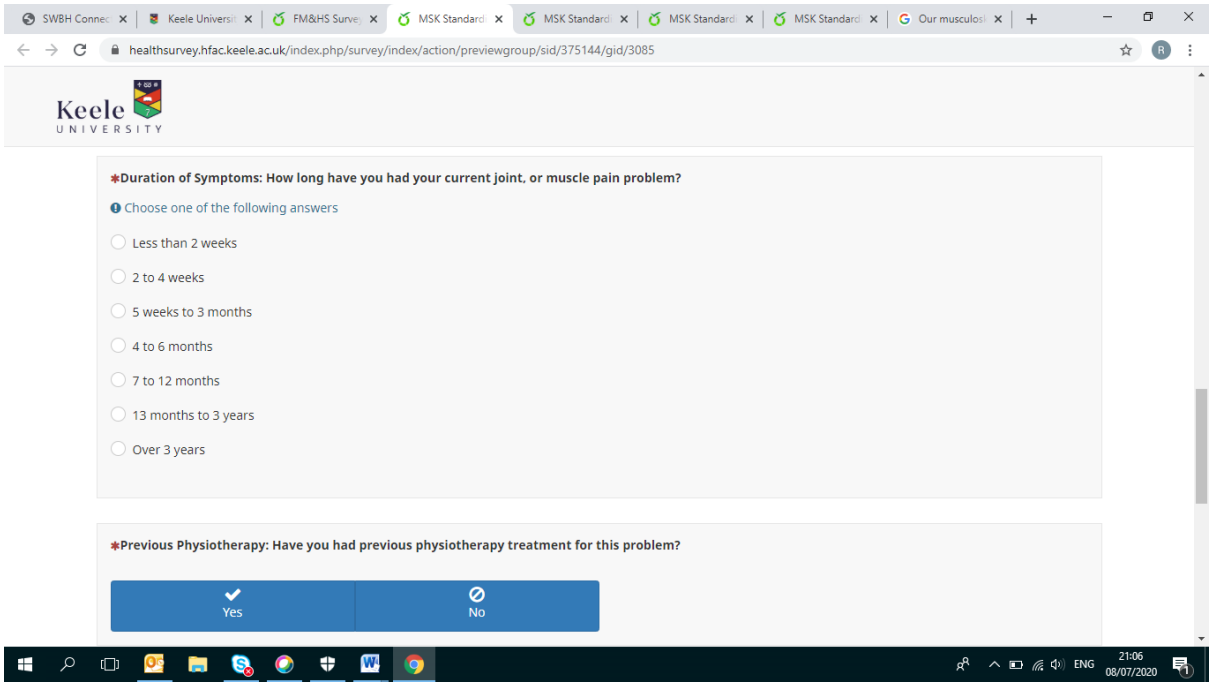
Section 1: About your condition:

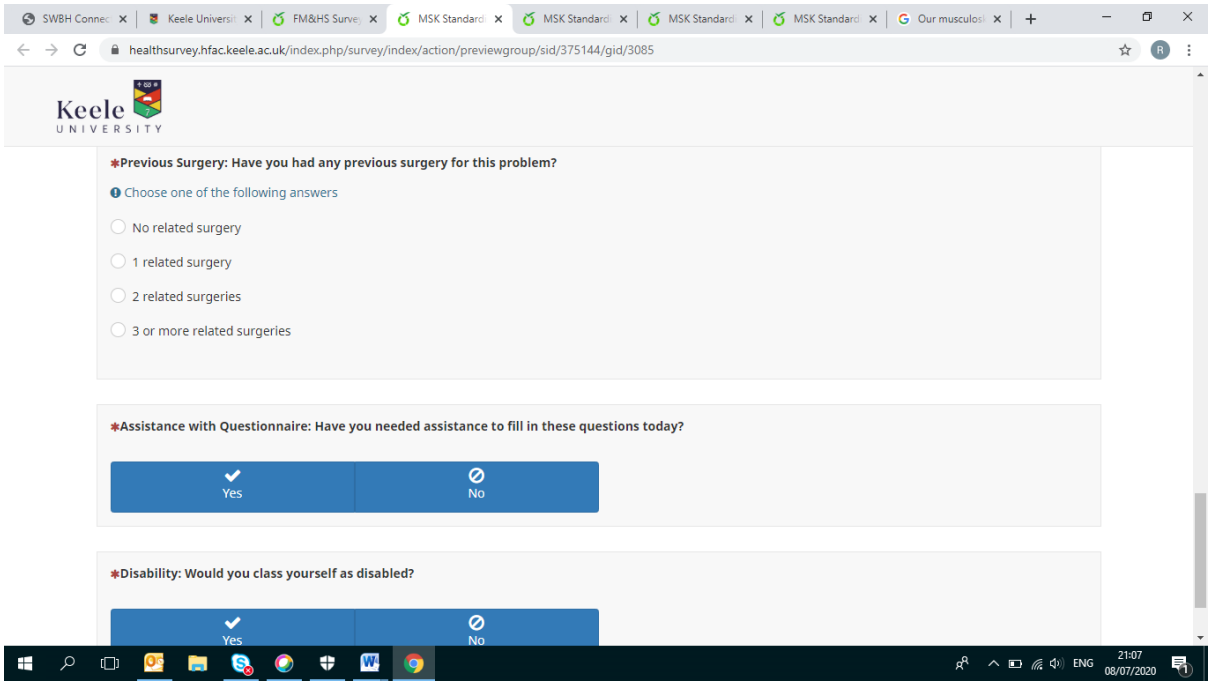
See content of demo in screenshots below (patients will be talked through this)

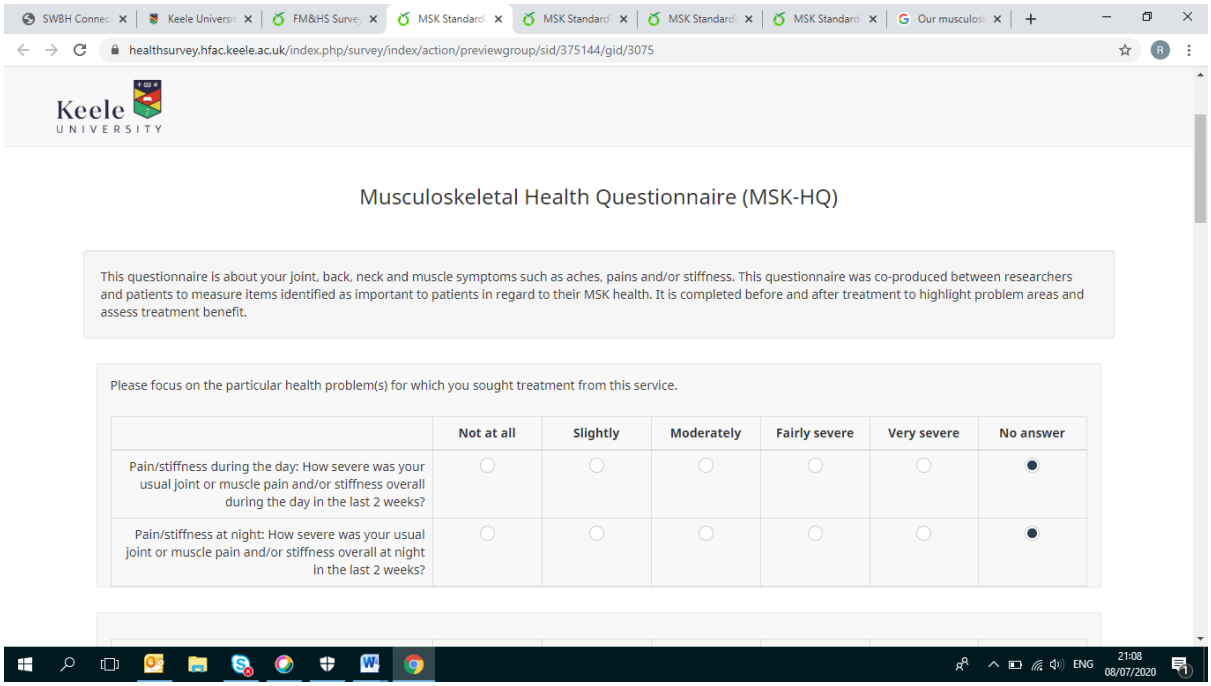


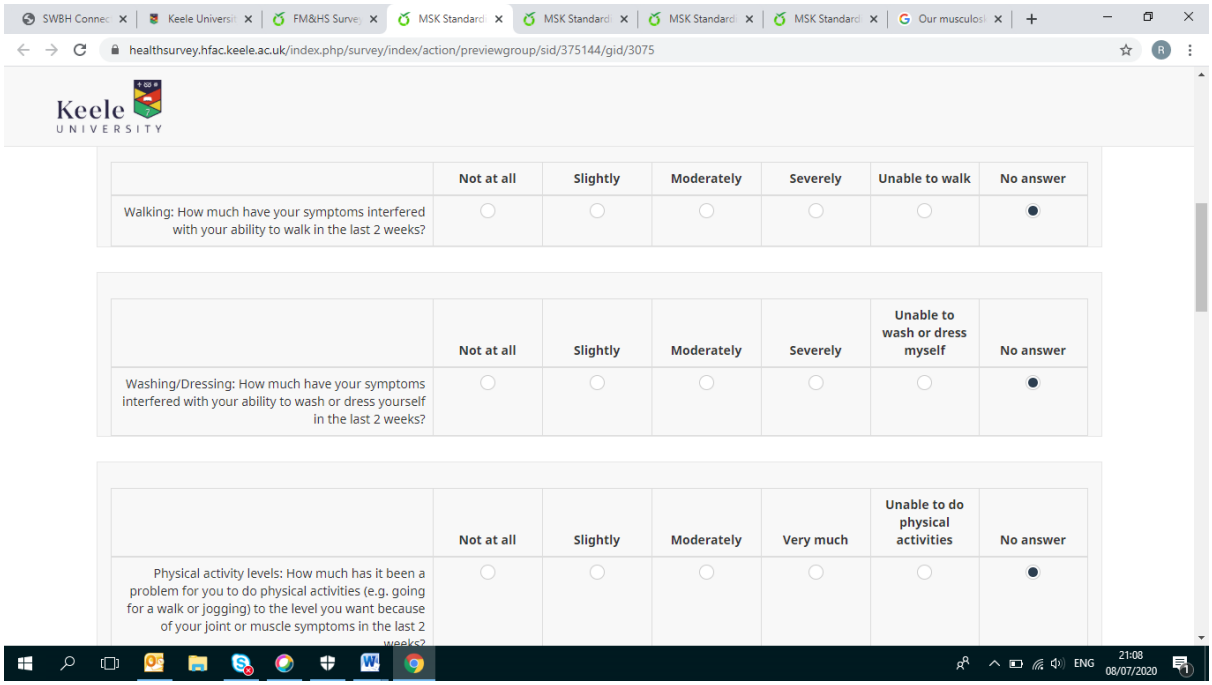



healthsurvey.hfac.keele.ac.uk/index.php/survey/index/action/previewgroup/sid/375144/gid/3085
Keele UNIVERSITY
*Other Health Conditions: Have you ever been diagnosed with any of the following conditions?
Check all that apply
Cancer
Chronic pain
Circulatory condition
Diabetes
Digestive system condition
Learning disability
Mental health condition
Arthritis
Neurological condition
Chronic kidney condition
Respiratory condition
None of the above
ENG
21:06
08/07/2020


healthsurvey.hfac.keele.ac.uk/index.php/survey/index/action/previewgroup/sid/375144/gid/3085
Keele UNIVERSITY
*Duration of Symptoms: How long have you had your current joint, or muscle pain problem?
Choose one of the following answers
Less than 2 weeks
2 to 4 weeks
5 weeks to 3 months
4 to 6 months
7 to 12 months
13 months to 3 years
Over 3 years
*Previous Physiotherapy: Have you had previous physiotherapy treatment for this problem?
Yes
No
ENG
21:06
08/07/2020


healthsurvey.hfac.keele.ac.uk/index.php/survey/index/action/previewgroup/sid/375144/gid/3085
Keele UNIVERSITY
*Previous Surgery: Have you had any previous surgery for this problem?
Choose one of the following answers
No related surgery
1 related surgery
2 related surgeries
3 or more related surgeries
*Assistance with Questionnaire: Have you needed assistance to fill in these questions today?
Yes
No
*Disability: Would you class yourself as disabled?
Yes
No
ENG
21:07
08/07/2020


healthsurvey.hfac.keele.ac.uk/index.php/survey/index/action/previewgroup/sid/375144/gid/3075
Keele UNIVERSITY
Musculoskeletal Health Questionnaire (MSK-HQ)
This questionnaire is about your joint, back, neck and muscle symptoms such as aches, pains and/or stiffness. This questionnaire was co-produced between researchers and patients to measure items identified as important to patients in regard to their MSK health. It is completed before and after treatment to highlight problem areas and assess treatment benefit.
Please focus on the particular health problem(s) for which you sought treatment from this service.
Not at all
Slightly
Moderately
Fairly severe
Very severe
No answer
Pain/stiffness during the day: How severe was your usual joint or muscle pain and/or stiffness overall during the day in the last 2 weeks?
Pain/stiffness at night: How severe was your usual joint or muscle pain and/or stiffness overall at night in the last 2 weeks?
ENG
21:08
08/07/2020



healthsurvey.hfac.keele.ac.uk/index.php/survey/index/action/previewgroup/sid/375144/gid/3075

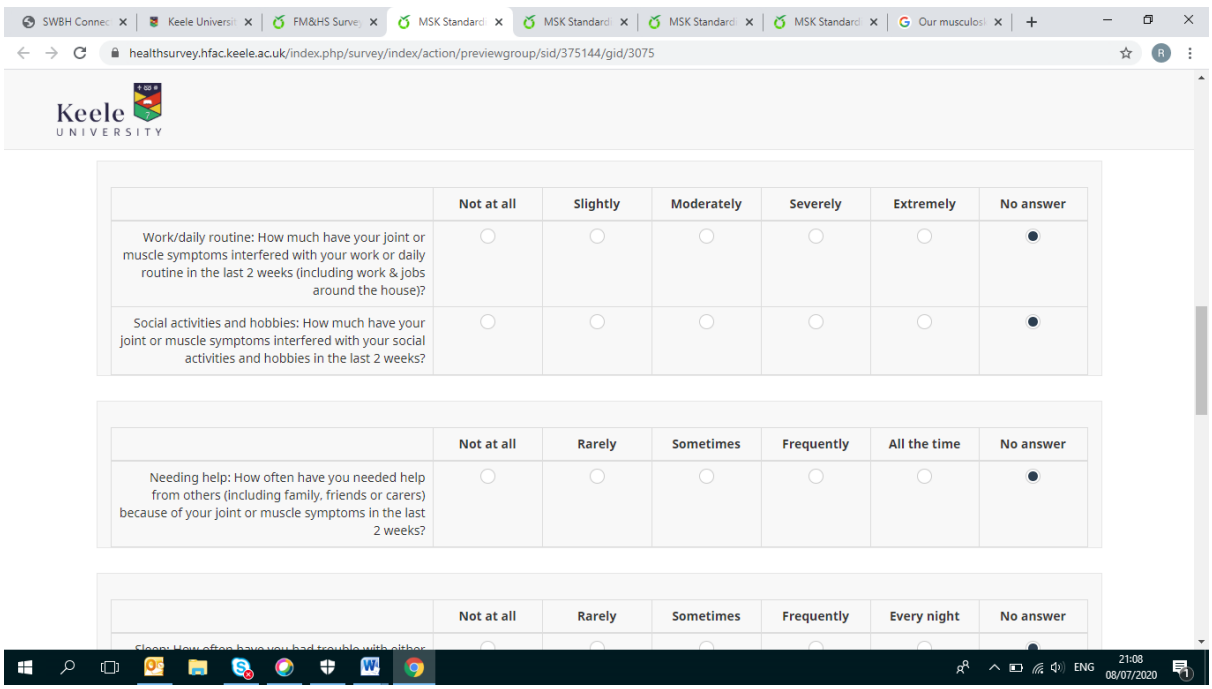
Keele UNIVERSITY

	Not at all	Slightly	Moderately	Severely	Unable to walk	No answer
Walking: How much have your symptoms interfered with your ability to walk in the last 2 weeks?	○	○	○	○	○	●

	Not at all	Slightly	Moderately	Severely	Unable to wash or dress myself	No answer
Washing/Dressing: How much have your symptoms interfered with your ability to wash or dress yourself in the last 2 weeks?	○	○	○	○	○	●

	Not at all	Slightly	Moderately	Very much	Unable to do physical activities	No answer
Physical activity levels: How much has it been a problem for you to do physical activities (e.g. going for a walk or jogging) to the level you want because of your joint or muscle symptoms in the last 2 weeks?	○	○	○	○	○	●

ENG 21:08 08/07/2020



healthsurvey.hfac.keele.ac.uk/index.php/survey/index/action/previewgroup/sid/375144/gid/3075

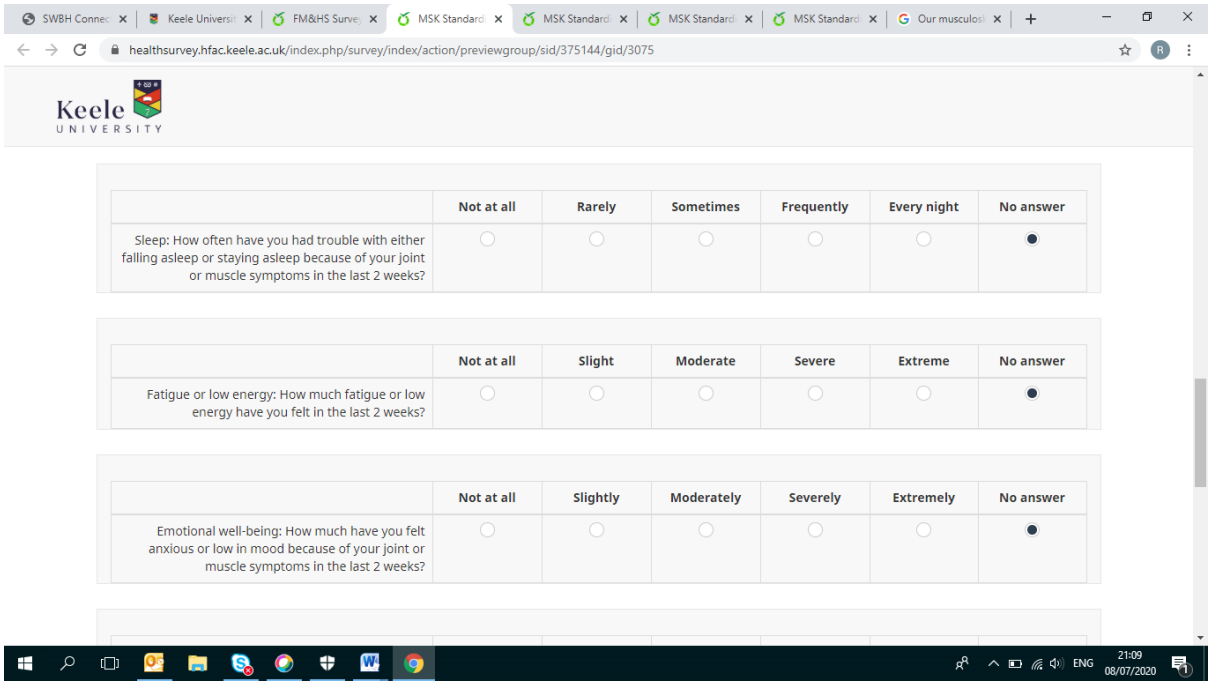
Keele UNIVERSITY

	Not at all	Slightly	Moderately	Severely	Extremely	No answer
Work/daily routine: How much have your joint or muscle symptoms interfered with your work or daily routine in the last 2 weeks (including work & jobs around the house)?	○	○	○	○	○	●
Social activities and hobbies: How much have your joint or muscle symptoms interfered with your social activities and hobbies in the last 2 weeks?	○	○	○	○	○	●

	Not at all	Rarely	Sometimes	Frequently	All the time	No answer
Needing help: How often have you needed help from others (including family, friends or carers) because of your joint or muscle symptoms in the last 2 weeks?	○	○	○	○	○	●

	Not at all	Rarely	Sometimes	Frequently	Every night	No answer
Sleep: How often have you had trouble with either	○	○	○	○	○	●

ENG 21:08 08/07/2020



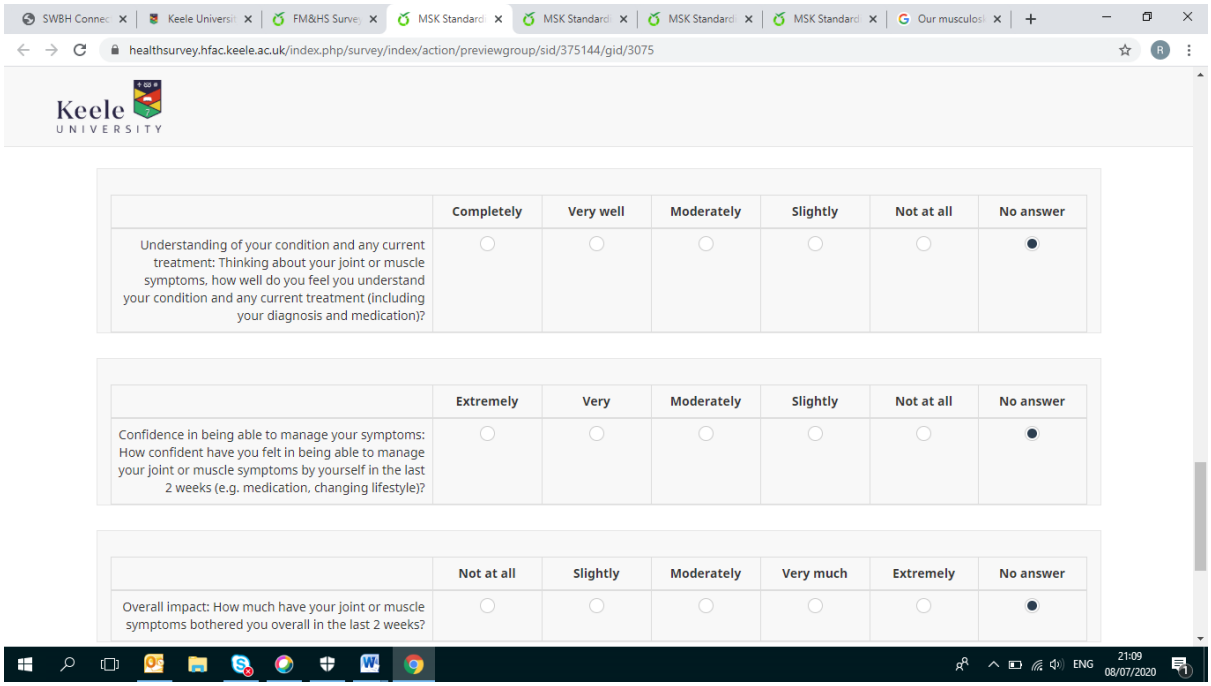
healthsurvey.hfac.keele.ac.uk/index.php/survey/index/action/previewgroup/sid/375144/gid/3075

Keele UNIVERSITY

	Not at all	Rarely	Sometimes	Frequently	Every night	No answer
Sleep: How often have you had trouble with either falling asleep or staying asleep because of your joint or muscle symptoms in the last 2 weeks?	○	○	○	○	○	●

	Not at all	Slight	Moderate	Severe	Extreme	No answer
Fatigue or low energy: How much fatigue or low energy have you felt in the last 2 weeks?	○	○	○	○	○	●

	Not at all	Slightly	Moderately	Severely	Extremely	No answer
Emotional well-being: How much have you felt anxious or low in mood because of your joint or muscle symptoms in the last 2 weeks?	○	○	○	○	○	●



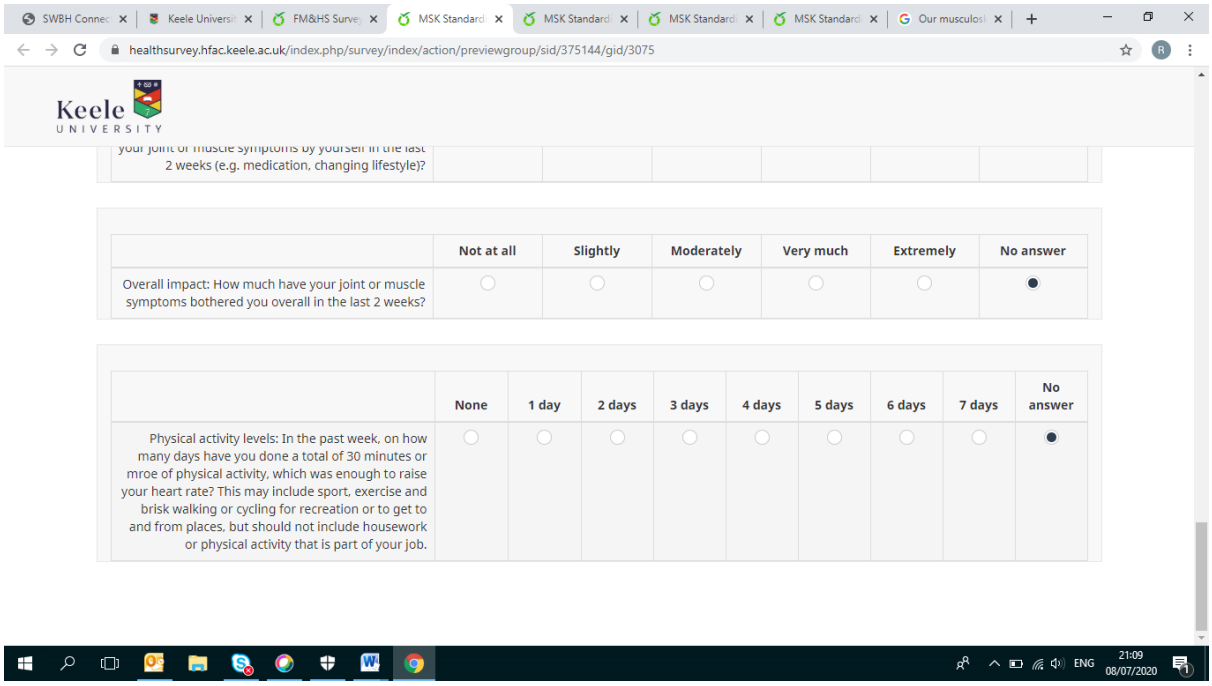
healthsurvey.hfac.keele.ac.uk/index.php/survey/index/action/previewgroup/sid/375144/gid/3075

Keele UNIVERSITY

	Completely	Very well	Moderately	Slightly	Not at all	No answer
Understanding of your condition and any current treatment: Thinking about your joint or muscle symptoms, how well do you feel you understand your condition and any current treatment (including your diagnosis and medication)?	○	○	○	○	○	●

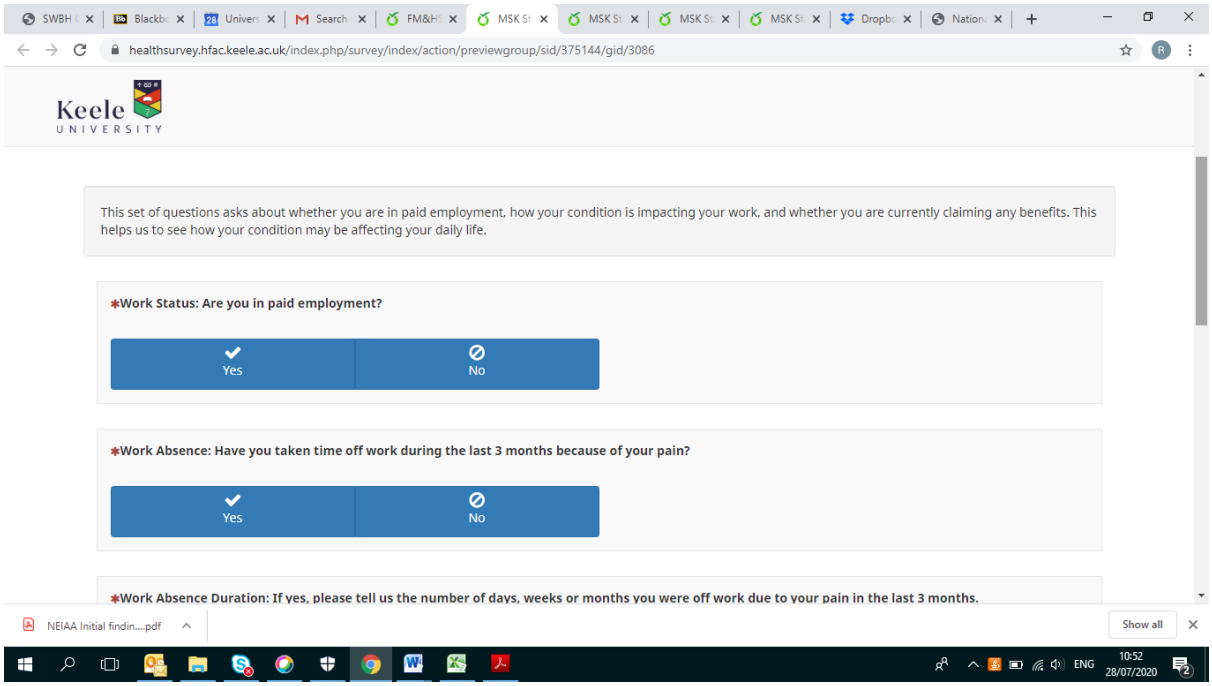
	Extremely	Very	Moderately	Slightly	Not at all	No answer
Confidence in being able to manage your symptoms: How confident have you felt in being able to manage your joint or muscle symptoms by yourself in the last 2 weeks (e.g. medication, changing lifestyle)?	○	○	○	○	○	●

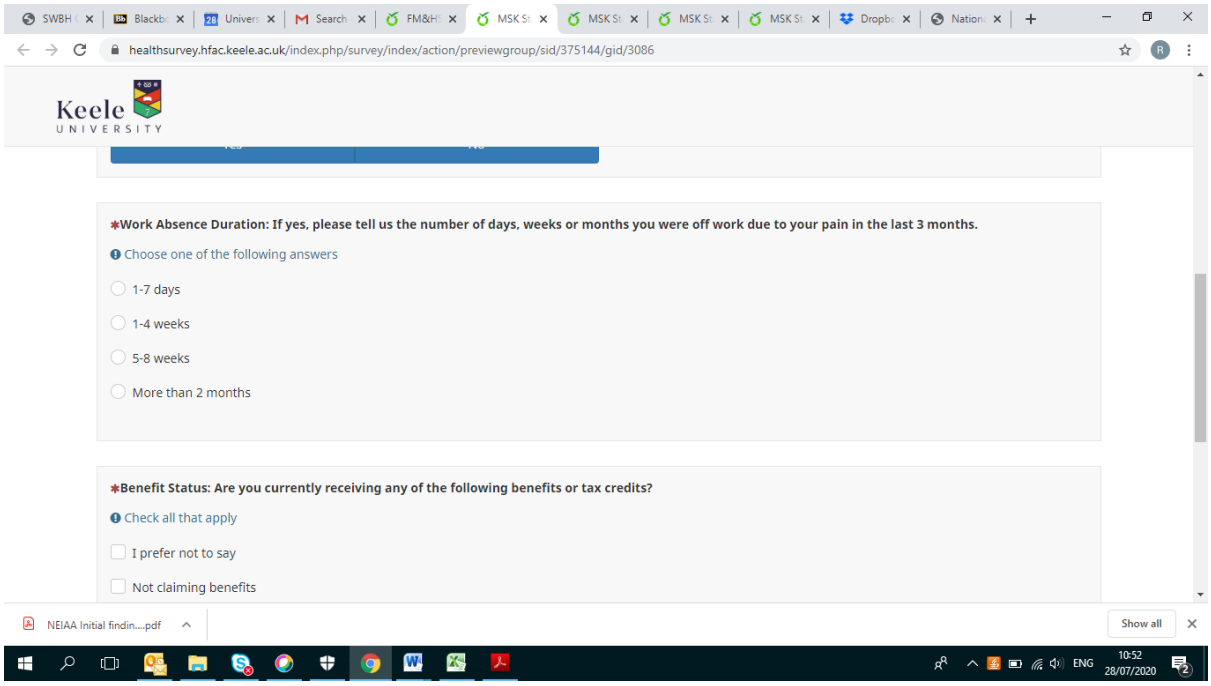
	Not at all	Slightly	Moderately	Very much	Extremely	No answer
Overall impact: How much have your joint or muscle symptoms bothered you overall in the last 2 weeks?	○	○	○	○	○	●

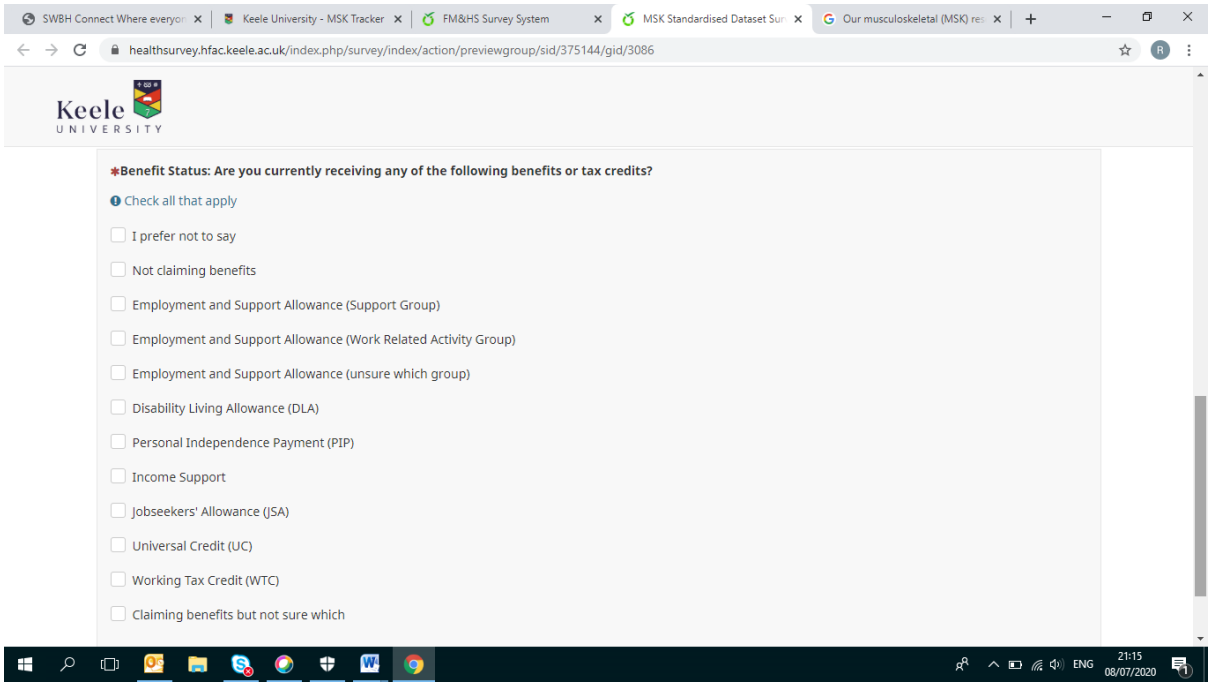


Section 2: About your work

See content of demo in screenshots below (patients will be talked through this)

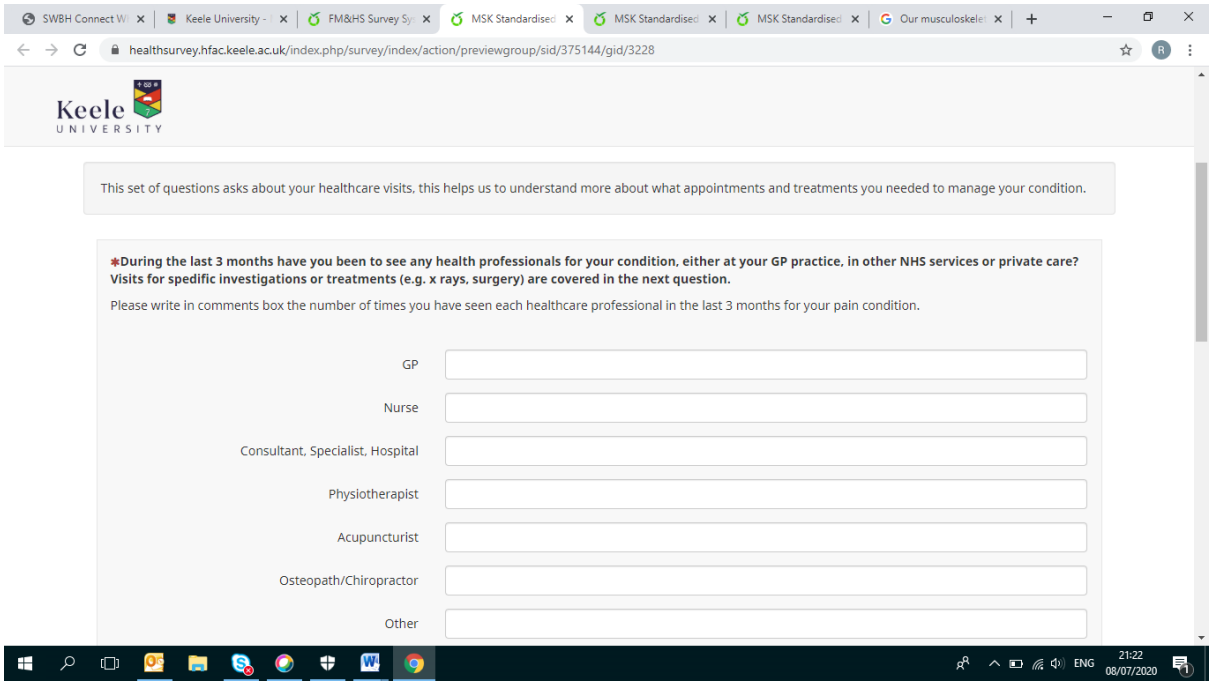


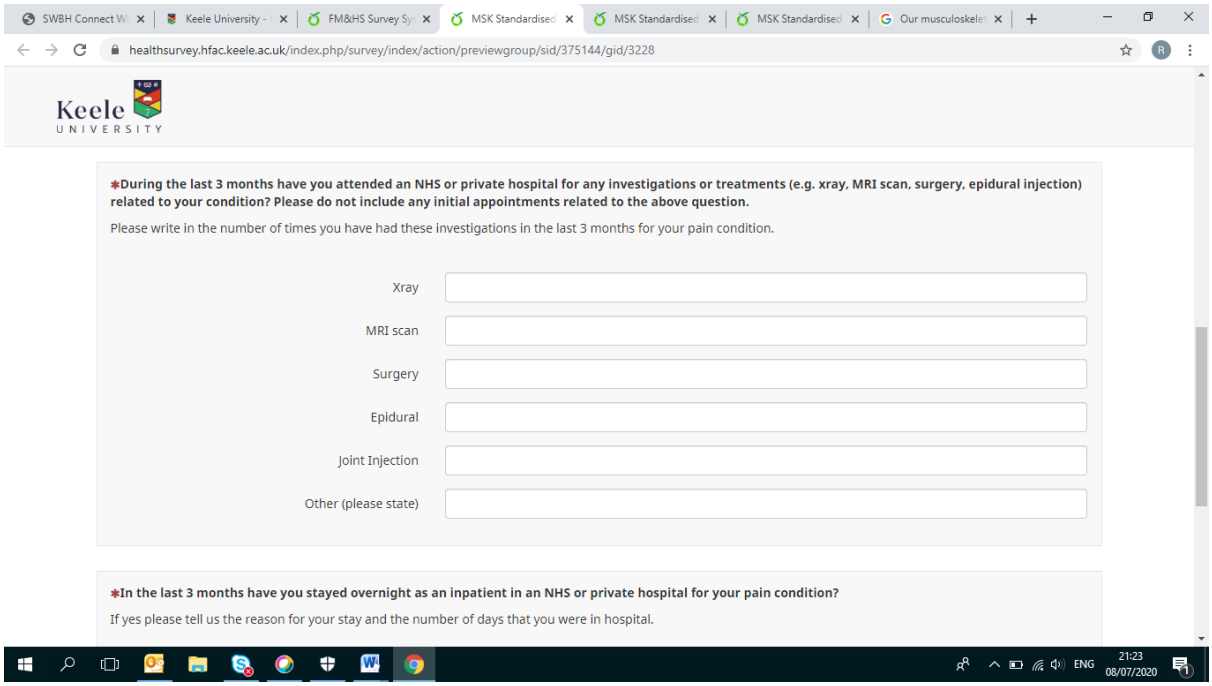


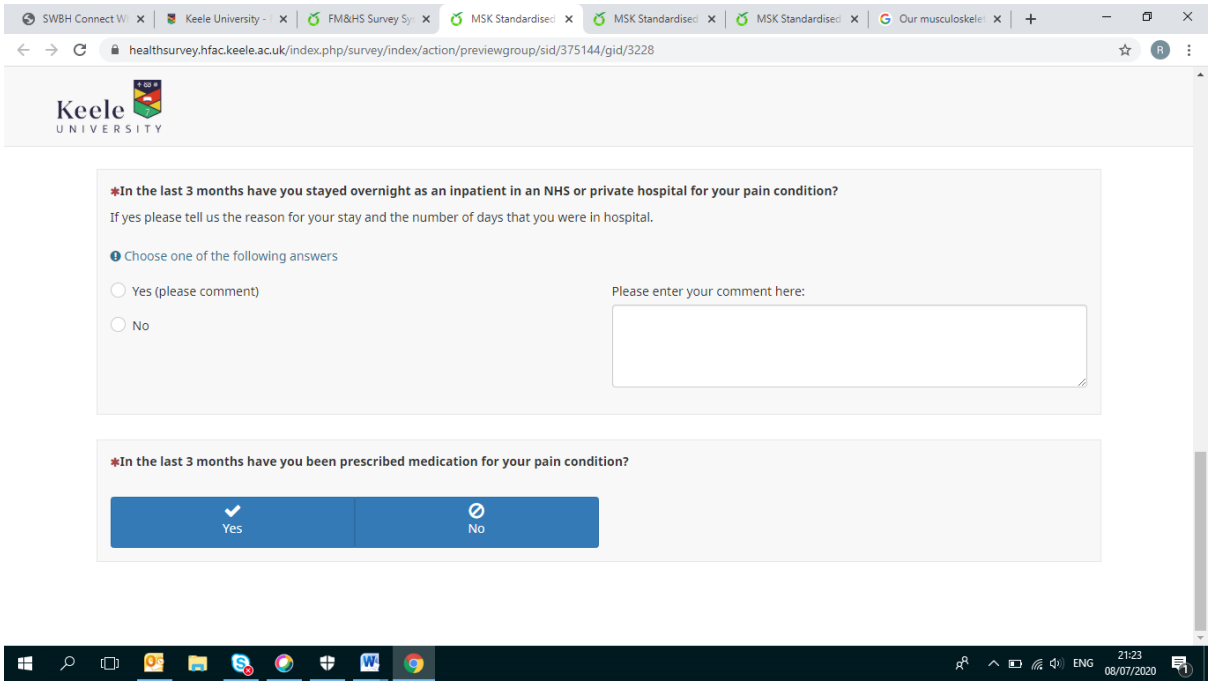


Section 3: About your healthcare visits

See content of demo in screenshots below (patients will be talked through this)

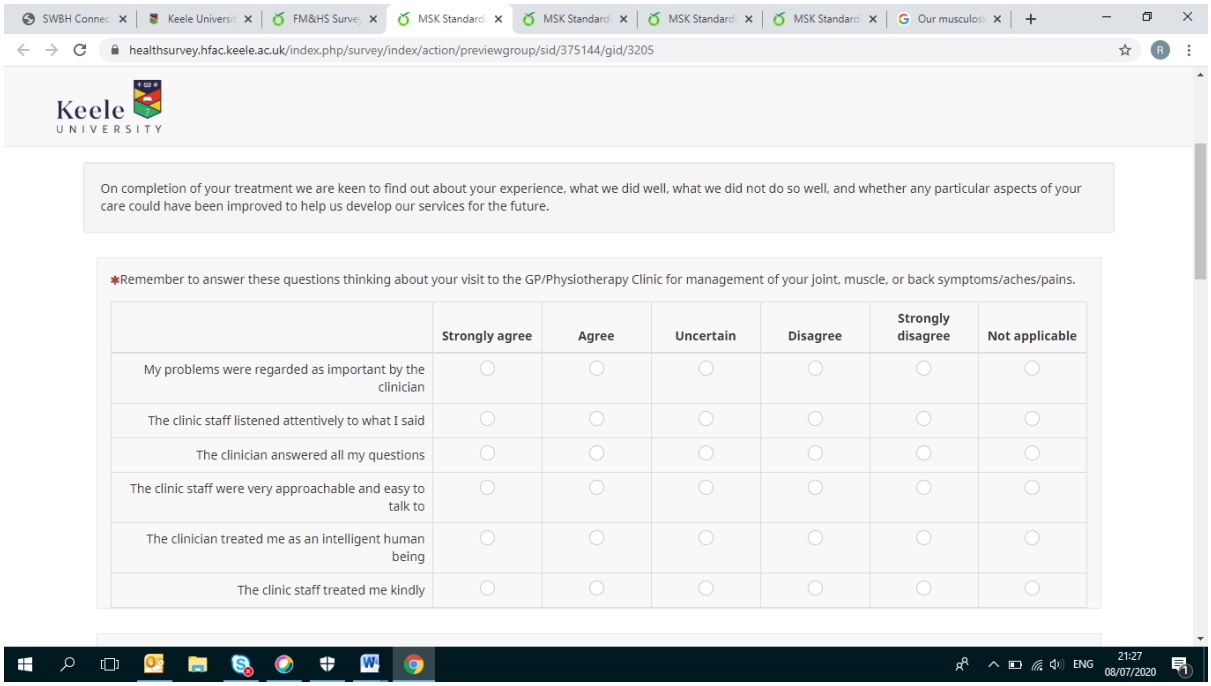

healthsurvey.hfac.keele.ac.uk/index.php/survey/index/action/previewgroup/sid/375144/gid/3228
Keele UNIVERSITY
This set of questions asks about your healthcare visits, this helps us to understand more about what appointments and treatments you needed to manage your condition.
*During the last 3 months have you been to see any health professionals for your condition, either at your GP practice, in other NHS services or private care? Visits for spedific investigations or treatments (e.g. x rays, surgery) are covered in the next question.
Please write in comments box the number of times you have seen each healthcare professional in the last 3 months for your pain condition.
GP
Nurse
Consultant, Specialist, Hospital
Physiotherapist
Acupuncturist
Osteopath/Chiropractor
Other
ENG
21:22
08/07/2020

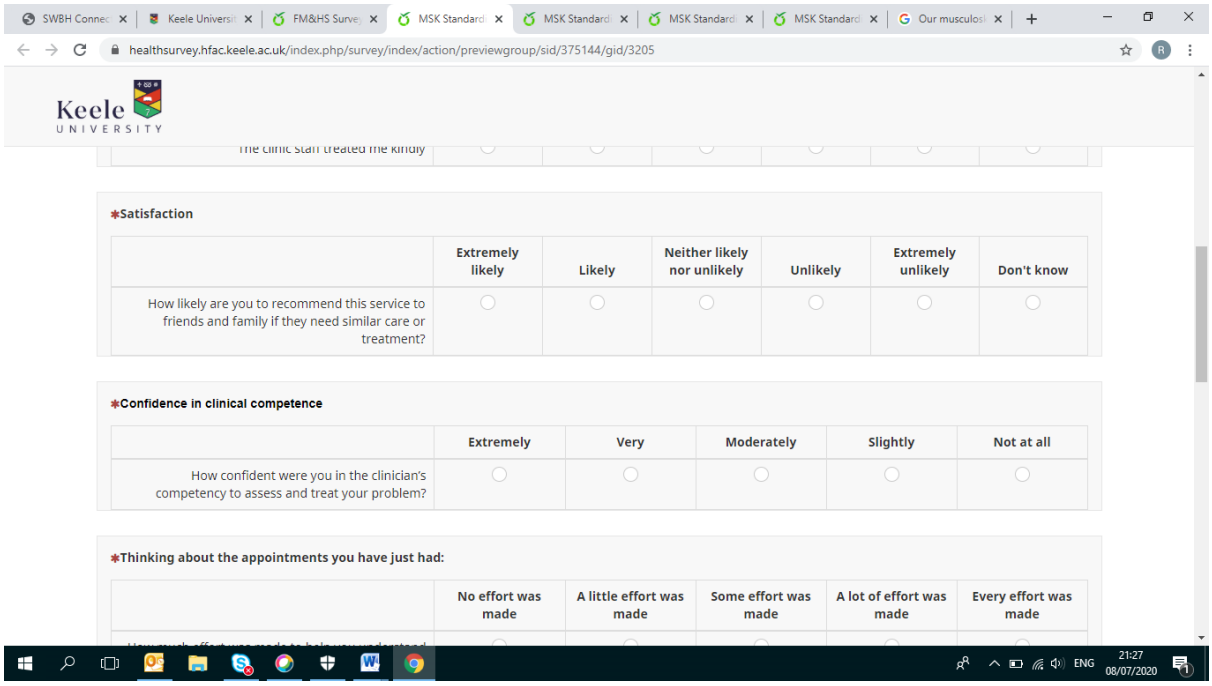

healthsurvey.hfac.keele.ac.uk/index.php/survey/index/action/previewgroup/sid/375144/gid/3228
Keele UNIVERSITY
*During the last 3 months have you attended an NHS or private hospital for any investigations or treatments (e.g. xray, MRI scan, surgery, epidural injection) related to your condition? Please do not include any initial appointments related to the above question.
Please write in the number of times you have had these investigations in the last 3 months for your pain condition.
Xray
MRI scan
Surgery
Epidural
Joint Injection
Other (please state)
*In the last 3 months have you stayed overnight as an inpatient in an NHS or private hospital for your pain condition?
If yes please tell us the reason for your stay and the number of days that you were in hospital.
ENG
21:23
08/07/2020

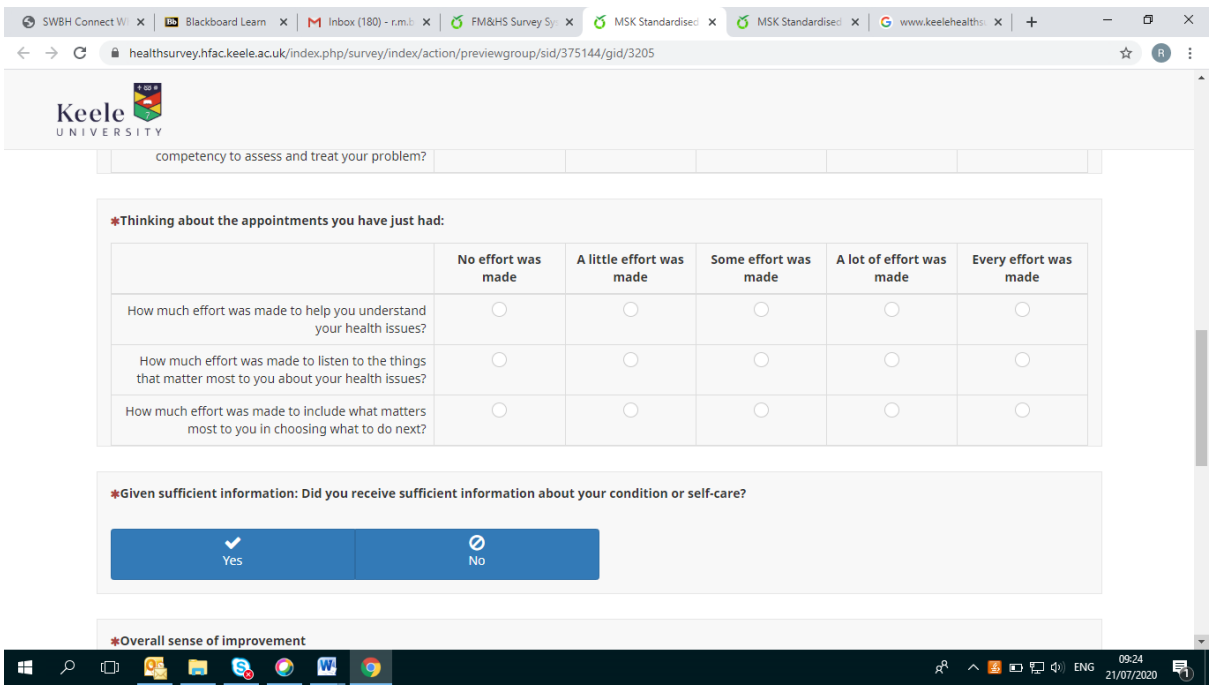


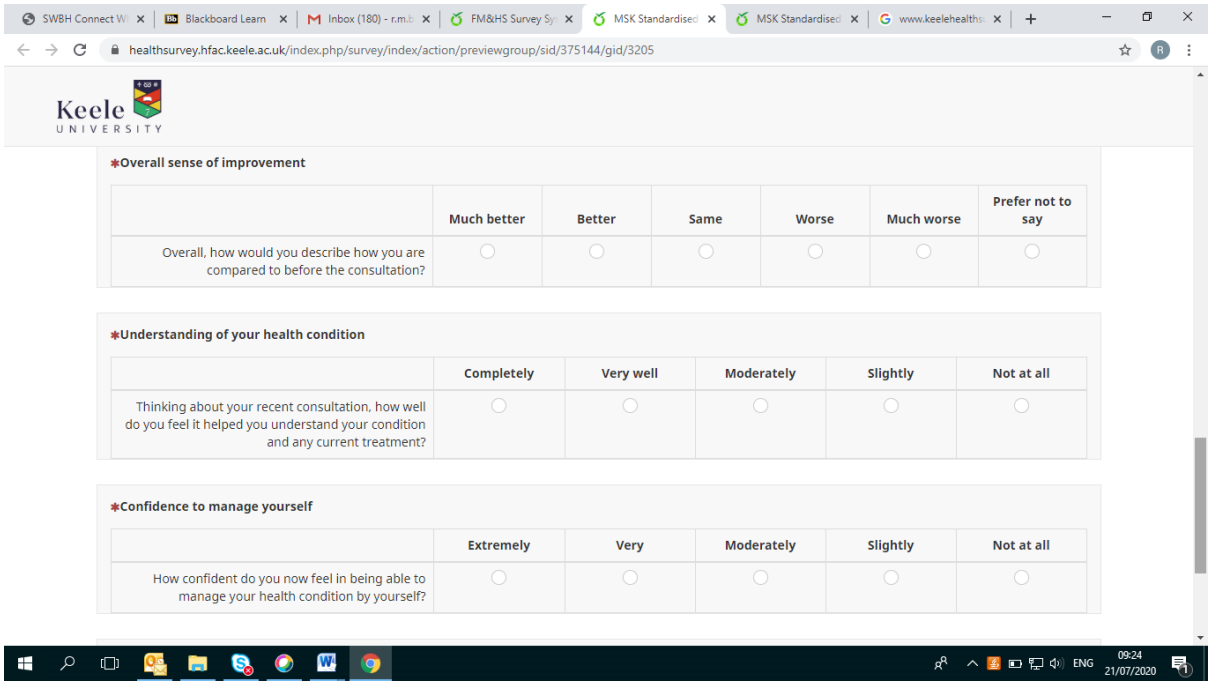
Section 4: About your experience

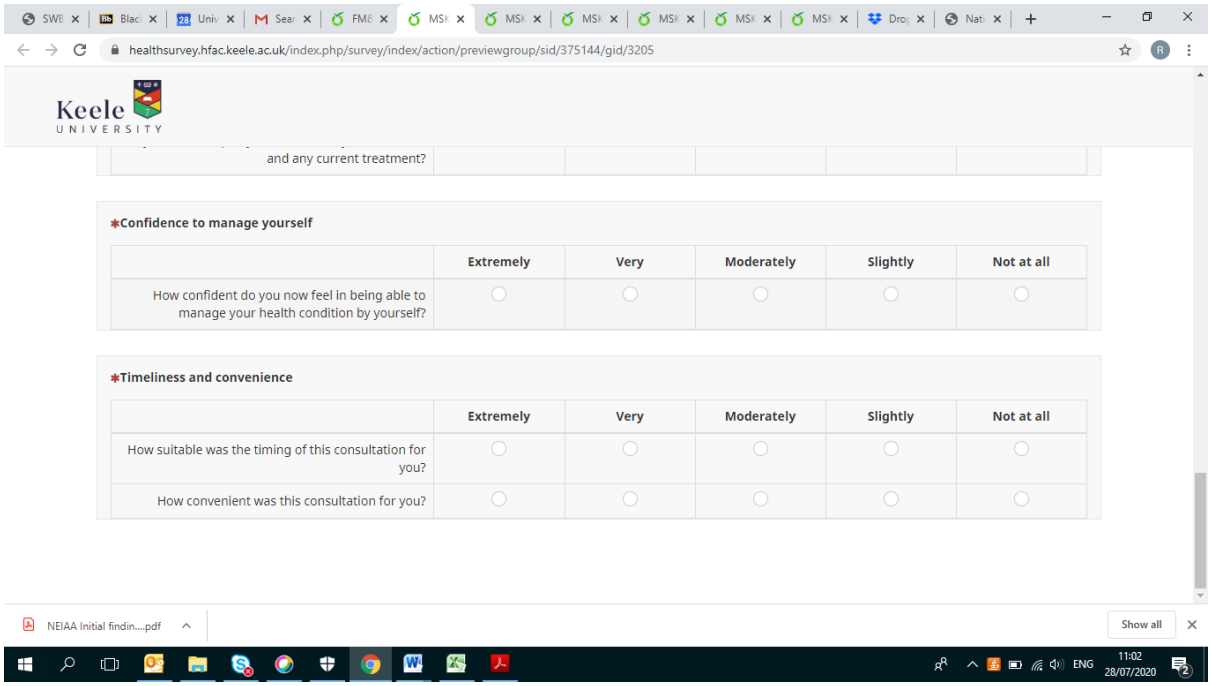
See content of demo in screenshots below (patients will be talked through this)




healthsurvey.hfac.keele.ac.uk/index.php/survey/index/action/previewgroup/sid/375144/gid/3205
Keele UNIVERSITY
The clinic staff treated me kindly
*Satisfaction
Extremely likely | Likely | Neither likely nor unlikely | Unlikely | Extremely unlikely | Don't know
How likely are you to recommend this service to friends and family if they need similar care or treatment?
*Confidence in clinical competence
Extremely | Very | Moderately | Slightly | Not at all
How confident were you in the clinician's competency to assess and treat your problem?
*Thinking about the appointments you have just had:
No effort was made | A little effort was made | Some effort was made | A lot of effort was made | Every effort was made
ENG 21:27 08/07/2020


healthsurvey.hfac.keele.ac.uk/index.php/survey/index/action/previewgroup/sid/375144/gid/3205
Keele UNIVERSITY
competency to assess and treat your problem?
*Thinking about the appointments you have just had:
No effort was made | A little effort was made | Some effort was made | A lot of effort was made | Every effort was made
How much effort was made to help you understand your health issues?
How much effort was made to listen to the things that matter most to you about your health issues?
How much effort was made to include what matters most to you in choosing what to do next?
*Given sufficient information: Did you receive sufficient information about your condition or self-care?
Yes
No
*Overall sense of improvement
ENG 09:24 21/07/2020


healthsurvey.hfac.keele.ac.uk/index.php/survey/index/action/previewgroup/sid/375144/gid/3205
Keele UNIVERSITY
*Overall sense of improvement
Much better | Better | Same | Worse | Much worse | Prefer not to say
Overall, how would you describe how you are compared to before the consultation?
*Understanding of your health condition
Completely | Very well | Moderately | Slightly | Not at all
Thinking about your recent consultation, how well do you feel it helped you understand your condition and any current treatment?
*Confidence to manage yourself
Extremely | Very | Moderately | Slightly | Not at all
How confident do you now feel in being able to manage your health condition by yourself?


healthsurvey.hfac.keele.ac.uk/index.php/survey/index/action/previewgroup/sid/375144/gid/3205
Keele UNIVERSITY
and any current treatment?
*Confidence to manage yourself
Extremely | Very | Moderately | Slightly | Not at all
How confident do you now feel in being able to manage your health condition by yourself?
*Timeliness and convenience
Extremely | Very | Moderately | Slightly | Not at all
How suitable was the timing of this consultation for you?
How convenient was this consultation for you?
NEIAA Initial findin....pdf
Show all

Appendix 5-2: Consensus study ethical approval letter

Keele University FMHS Faculty Research Ethics Committee
health.ethics@keele.ac.uk

11th September 2020

Dear Roanna

Project Title:	Standardised Outcomes for MSK Services: seeking consensus using an online survey.
REC Project Reference:	MH-200141
Type of Application	Amendment

Keele University's Faculty of Medicine and Health Sciences Research Ethics Committee (FMHS FREC) reviewed the above amendment.

Favourable Ethical opinion

The members of the Committee gave a favourable ethical opinion of the above research on the basis described in the application form, protocol and supporting documentation, subject to the conditions specified below.

Conditions of the favourable opinion

The favourable opinion is subject to the following conditions being met prior to the implementation of the amendment.

1.	None

Reporting requirements

The University's standard operating procedures give detailed guidance on reporting requirements for studies with a favourable opinion including:

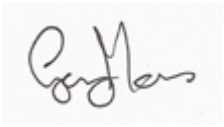
- Notifying substantial amendments
- Notifying issues which may have an impact upon ethical opinion of the study
- Progress reports
- Notifying the end of the study

Approved documents

The documents reviewed and approved are:

Document	Version	Date
All documents submitted with MH-200141		

Yours sincerely,



Dr Gary Moss

Chair

Appendix 5-3 Consensus study supporting information

MSK Community Services Standardised Dataset

Introduction

The NHS Mandate (2018) lays out the need for NHS transformation, with NHS England supporting leaders to drive forwards real improvements in patient care and patient outcomes. Tackling unwarranted variation is highlighted as a priority objective within both the NHS Mandate (2018) and the Five Year Forward View (NHS England, 2014), aiming to reduce the 'unacceptable' care and quality gap.

Standardised data is essential in order to identify variation in Musculoskeletal (MSK) service performance (including outcomes and costs) and requires the use of specific standardised metrics. There has been a large focus on costing, efficiency, and standardised metrics within the acute MSK setting, but far less attention in primary care and community services. In response to the COVID-19 pandemic there is also increasing focus on MSK digital health tools, but evaluation of these innovations is made difficult by the large number of outcome measures used in musculoskeletal conditions which makes comparing different models of care challenging (Hewitt et al, 2020).

Keele Primary Care Centre Versus Arthritis have therefore developed an evidence-based set of core metrics that make up a recommended standardised dataset to be used by UK community and primary care MSK services. This document outlines the proposed metrics and tools included within the dataset, with supporting detail for implementation.

The dataset is made up of core areas of; demographic factors, clinical factors, employment factors, functional/MSK health status, patient reported experience measures, and healthcare utilisation (economic factors). This is a collection of evidence based validated tools such as the Musculoskeletal Health Questionnaire (MSK-HQ) (Hill et al, 2016), and patient factors/metrics including demographics and characteristics that can be used for case-mix adjustment (a statistical process that aims to account for differences in the mix of patient attributes/characteristics across definitive patient cohorts (Iezzoni, 2009)) in order to be able to make objective comparisons of PROM data (Deutscher et al, 2018)). Where there is overlap, factors have been aligned with those of ICHOM to improve global standardisation (ICHOM 2017). Factors including specific questions and coding are listed within an accompanying excel document, with more detail on included tools/variables outlined below. All mandatory tools included are free to use subject to obtaining the associated licence agreements (as shown below).

This MSK standardised dataset is currently in consultation phase. Over the next 12 months further data analysis will be undertaken to verify appropriate case-mix adjustment variables and to make recommendations on the most parsimonious case-mix adjustment model to be used within this setting. Feedback will also be collected from clinicians, service managers and patients looking to gain consensus over the core metrics to be included within the final published dataset.

Proposed Mandatory Variables within the Dataset

Variable Name	Response Options	Capture Point
Demographics		
Age	Continuous numeric	Baseline
Sex at birth	Binary (male/female)	Baseline
Education	Categorical (4 options)	Baseline
Ethnicity	Categorical (5 options)	Baseline
Baseline Clinical Factors		
Pain Site	Categorical (11 options)	Baseline
Comorbidities	Categorical (12 options)	Baseline
Duration of Symptoms	Categorical (5 options)	Baseline
Previous Surgery	Categorical (4 options)	Baseline
Self-Reported as Disabled	Binary (yes/no)	Baseline
Employment		
Work Status	Binary (yes/no)	Baseline and 3 months
Work Absence	Binary (yes/no)	Baseline and 3 months
Work Absence Duration	Categorical (4 options)	Baseline and 3 months
Functional Status		
MSK-HQ (MSK Health Status)	Questionnaire (15 questions)	Baseline and 3 months
Pain Intensity (NPRS)	Numeric (0-10)	Baseline and 3 months
Patient Reported Experience		
Friends and Family Test (FFT)	Questionnaire (2 questions)	3 months
Global Change in Health Status	Categorical (6 options)	3 months

Proposed Optional Variables

Variable Name	Response Options	Capture Point
Baseline Clinical Factors		
Previous Physiotherapy	Binary (yes/no)	Baseline
Assisted with Questionnaire	Binary (yes/no)	Baseline
Employment		
Benefit Status	Categorical (12 options)	Baseline
Functional Status		
STarT MSK (Risk Status)	Questionnaire (10 questions)	Baseline
EQ5D5L (QOL)	Questionnaire (5 questions)	Baseline and 3 months
Patient Reported Experience		
Valuing Patients as Individuals -Care and Respect -Understanding & Engagement	Questionnaire (6 questions)	3 months
CollaboRATE -Shared Decision making	Questionnaire (3 questions)	3 months
MSK Indicators -Clinical Competence -Sufficient Information	2 questions	3 months
Economic Factors		
Healthcare Utilisation	Free text numeric	3 months
Investigations and Treatments	Free text numeric	3 months
Inpatient Stays	Free text numeric	3 months

Prescribed Medication	Binary (yes/no)	3 months

Demographic Factors

Age: (continuous numeric) Year of birth used to ensure patients are not identifiable from the anonymised data. This variable is for both use in case-mix adjustment and descriptive analysis.

Sex: (at birth) (binary) Research does not support the use of gender as a case-mix adjuster (Burgess et al, 2019, 2020) but this is still important for descriptive analysis and is included in similar core datasets in this area (ICHOM, 2017, Clement et al, 2015, Rolfson et al, 2016).

Education: (categorical (4 options)) The Education variable is for case-mix adjustment of the data as a proxy for socioeconomic status, again to ensure anonymity of patients rather than use of postcode/Index of Multiple Deprivation (IMD). This is supported by the latest Focus on Therapeutic Outcomes (FOTO) case-mix adjustment model (Deutscher et al, 2018) and by ICHOM (2017) and categories are aligned to ICHOM (2017).

Ethnicity: (categorical (5 options)) The Ethnicity variable can be used within case-mix adjustment modelling as supported by National PROMs (NHS England, 2013) but is also necessary for descriptive analysis to highlight variation across groups. Groups are informed by the Office of National Statistics (ONS, 2019).

Baseline Clinical Factors

MSK Pain Site: This is a list of 11 potential pain sites, patients mark all as appropriate (yes/no pain) to list all problematic pain sites (This variable is for use in case-mix adjustment and for descriptive analysis allowing for targeted quality improvement).

Comorbidities: This is a list of 12 comorbid conditions, patients mark all as appropriate (yes/no) to list all comorbid conditions (This variable is informed by the new NICE Indicator for multi-morbidity in primary care (NICE, 2019) and forms part of the case-mix adjustment model alongside allowing for descriptive analysis around complexity).

Duration of Symptoms: (categorical (5 options)) This variable forms part of the case-mix adjustment model and will help highlight differences between services compared with regards to case-mix and chronicity of population.

Previous Surgery: (categorical (4 options)) This variable aligns to the latest FOTO case-mix adjustment model (Deutscher et al, 2018) and is for use in case-mix adjustment.

Self-Reported as Disabled: (categorical (2 options)) This variable aligns to the National PROMs Programme case-mix adjustment model (NHS England, 2013) and is for use in case-mix adjustment.

Optional Extras:

Previous Physiotherapy: This variable will be tested within the case-mix model to see if it adds to predictive ability but is also useful for signposting of patents that have had previous treatment.

Assisted with Questionnaire (Q1): This variable is aligned to the National PROMs Programme case-mix adjustment model (NHS England, 2013), and will be further tested in planned data analysis. Within the National PROMs model assistance filling out the questionnaire at Q2 (follow up questionnaire) was predictive but not at Q1 (baseline questionnaire). For this model all variables will be collected at baseline for case-mix adjustment therefore giving less support for its inclusion.

Employment

Work Status (categorical (2 options))

Work Absence (categorical (2 options)) (only complete if at work)

Work Absence Duration (continuous numeric) (only complete if have been absent from work)

These employment factors are captured at baseline and are reflective of factors captured within the First Contact Physiotherapist Pilot commissioned by NHS England (funded by the Chartered Society of Physiotherapy Charitable Trust and the Department for Work and Pensions/DH Joint Work and Health Unit. Evaluation led by Keele's Primary Care Centre Versus Arthritis in collaboration with Nottingham University (CSP, 2018)).

Optional Extras

Benefit Status: (categorical (12 options)) This variable is similar to the 'payer' variable within the FOTO case-mix adjustment model (Deutscher et al, 2018) as it could be used as a proxy for socioeconomic status, it only needs to be asked of those patients reporting that they are not in paid employment.

Functional Status/MSK Health Status

MSK-HQ: (14 questions that make up MSK-HQ score (low (0) to high functioning (56) and 1 additional standalone question on Physical Activity Level) The MSK-HQ questionnaire was developed in 2016 (Hill et al 2016) as an MSK specific PROM for generic use across MSK conditions to measure patient's MSK health and response to MSK treatments. Over 300 licences have now been issued demonstrating good uptake across the MSK community. Additional information on development and scoring are available on the Oxford Innovations website where

a free licence can also be obtained; see https://innovation.ox.ac.uk/outcome-measures/musculoskeletal-health-questionnaire-msk-hq/

Optional Extras

STarT MSK: This is a baseline risk stratification tool and is made up of 10 questions including pain intensity using a numeric pain rating scale (NPRS). The STarT MSK is a risk stratification tool that places patients into categories dependent on their risk of a poor outcome (low, medium and high) (Campbell et al, 2016). Additional information can be found on the Keele website where a free licence can be obtained; see: https://www.keele.ac.uk/startmsk/ The full trial for the STarT MSK is still underway so although this is available for use and is useful for risk stratification it will not form part of the core recommended set until supporting evidence is available from the main trial alongside supporting matched treatment approaches.

EQ5D: The EQ-5D is a measure of quality of life and is an important tool for economic evaluation (Euroqol, 2019). The EQ-5D requires a licence agreement and is not free for all users and therefore has not been included at this stage in the core set. It is however licensed within NHS Secondary Care Trusts as part of the National PROMs Programme (NHS England, 2017) which uses it within all mandated data collections and therefore holds a licence agreement across NHS secondary care providers. The EQ5D will be further tested as a useful addition to the dataset within planned analysis.

Patient Reported Experience

Friends and Family Test (FFT): (made up of 2 questions, 1 with free text) The Friends and Family Test (FFT) was launched in 2013 and is now used by most NHS services (NHS England 2019). More information on implementation of this tool can be found in the FFT Guidance Document developed by NHS England: https://assets.nhs.uk/prod/documents/FFTGuide_Final_1807_FINAL.pdf and on the NHS England website: https://www.nhs.uk/using-the-nhs/about-the-nhs/friends-and-family-test-fft/

Global Change: (Categorical (6 options)) Global change is a useful measure of change in health and can be used alongside other PROMs to evaluate efficacy of intervention/pathway of care

Optional Extras

Valuing Patients as Individuals Scale (VPAI): (Only questions 3,7,9, for Care and Respect, and 1,4,8, for Understanding and Engagement, were included as others are not relevant to the community/primary care setting) The Valuing Patients as Individuals Scale (VPAI) was developed by Coyle and Williams in 2001. It is free to use and more details about the tool can be found in the paper by Jones et al (2017): DOI: 10.1111/jocn.13845, http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1365-2702/earlyview

CollaboRATE: The collaboRATE tool was developed by Elwyn et al (2013) to measure shared decision making as part of a clinical encounter. It is made up of 3 brief questions, has been designed for use in routine practice, and has undergone psychometric testing (Barr et al, 2014).

MSK Indicators: Additional indicators were used alongside validated patient reported experience measures (PREMs) including the VPAI and FFT to capture patient experience within the National FCP Evaluation (see Appendix 1 for details) which we recommend are adopted nationally.

Keele's Recommended List of Patient Experience Questions (See Appendix 5-3-1 for full list)

Economic Factors

Optional Extras

Healthcare Utilisation: (7 categories with patients annotating number of visits in past 3 months (to GP, Physiotherapist, Consultant etc))

Investigations and Treatments: (free text to enter investigation and frequency)

Inpatient Stays: (free text to enter reason for admission and length of stay)

Prescribed Medication: (binary yes/no for being prescribed medication for current pain condition)

The above economic factors are supported by a recently published systematic review (Burgess et al, 2020). This review found that the key drivers of MSK healthcare costs were GP visits, Outpatient Medical Specialist visits and Physiotherapy visits, followed by prescription medication and inpatient stays. Investigations including Xray and MRI, and private healthcare visits were other useful additions.

References

Barr, P.J., Thompson, R., Walsh, T., Grande, S.W., Ozanne, E.M. and Elwyn, G., 2014. The psychometric properties of CollaboRATE: a fast and frugal patient-reported measure of the shared decision-making process. *Journal of medical Internet research*, *16*(1), p.e2.

Burgess, R., Bishop, A., Lewis, M. and Hill, J., 2019. Models used for case-mix adjustment of patient reported outcome measures (PROMs) in musculoskeletal healthcare: A systematic review of the literature. *Physiotherapy*, *105*(2), pp.137-146.

Burgess, R., Mansell, G., Bishop, A., Lewis, M. and Hill, J., 2020. Predictors of functional outcome in musculoskeletal healthcare: An umbrella review. *European Journal of Pain*, *24*(1), pp.51-70.

Burgess, R., Hall, J., Bishop, A., Lewis, M. and Hill, J., 2020. Costing Methodology and Key Drivers of Health Care Costs Within Economic Analyses in Musculoskeletal Community and Primary Care Services: A Systematic Review of the Literature. *Journal of Primary Care & Community Health*, *11*, p.2150132719899763.

Campbell, P., Hill, J.C., Protheroe, J., Afolabi, E.K., Lewis, M., Beardmore, R., Hay, E.M., Mallen, C.D., Bartlam, B., Saunders, B. and van der Windt, D.A., 2016. Keele Aches and Pains Study protocol: validity, acceptability, and feasibility of the Keele STarT MSK tool for subgrouping musculoskeletal patients in primary care. *Journal of pain research*, *9*, p.807.

Clement, R.C., Welander, A., Stowell, C., Cha, T.D., Chen, J.L., Davies, M., Fairbank, J.C., Foley, K.T., Gehrchen, M., Hagg, O. and Jacobs, W.C., 2015. A proposed set of metrics for standardized outcome reporting in the management of low back pain. *Acta orthopaedica*, *86*(5), pp.523-533.

Coyle, J. and Williams, B., 2001. Valuing people as individuals: development of an instrument through a survey of person-centredness in secondary care. *Journal of Advanced Nursing*, *36*(3), pp.450-459.

Chartered Society of Physiotherapy (2018) FCP Webinar 2018. Available at: https://www.csp.org.uk/system/files/documents/2018-10/phase_3_webinar_15_10_18.pdf Accessed 06/08/19

Department of Health and Social Care (2018) NHS Mandate 2017 to 2018. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/692140/NHSE_Mandate__2017-18_revised.pdf Accessed 06/08/19

Deutscher, D., Werneke, M.W., Hayes, D., Mioduski, J.E., Cook, K.F., Fritz, J.M., Woodhouse, L.J. and Stratford, P.W., 2018. Impact of Risk Adjustment on Provider Ranking for Patients With Low Back Pain Receiving Physical Therapy. *journal of orthopaedic & sports physical therapy*, *48*(8), pp.637-648.

Elwyn, G., Barr, P.J., Grande, S.W., Thompson, R., Walsh, T. and Ozanne, E.M., 2013. Developing CollaboRATE: a fast and frugal patient-reported measure of shared decision making in clinical encounters. *Patient education and counseling*, *93*(1), pp.102-107.

Euroqol. 2019. EQ-5d-5L User Guide. Available at: https://euroqol.org/wp-content/uploads/2019/09/EQ-5D-5L-English-User-Guide_version-3.0-Sept-2019-secured.pdf Accessed 05/11/19

Hewitt, S., Sephton, R. and Yeowell, G., 2020. The Effectiveness of Digital Health Interventions in the Management of Musculoskeletal Conditions: Systematic Literature Review. *Journal of Medical Internet Research*, *22*(6), p.e15617.

Hill, J.C., Kang, S., Benedetto, E., Myers, H., Blackburn, S., Smith, S., Dunn, K.M., Hay, E., Rees, J., Beard, D. and Glyn-Jones, S., 2016. Development and initial cohort validation of the Arthritis Research UK Musculoskeletal Health Questionnaire (MSK-HQ) for use across musculoskeletal care pathways. *BMJ open*, *6*(8), p.e012331.

ICHOM 2017. ICHOM Low Back Pain Data Collection Reference Guide. Available at: https://ichom.org/files/medical-conditions/low-back-pain/low-back-pain-reference-guide.pdf Accessed 06/08/19

ICHOM 2017. ICHOM Hip and Knee Osteoarthritis Data Collection Reference Guide. Available at: https://ichom.org/files/medical-conditions/hip-knee-osteoarthritis/hip-knee-osteoarthritis-reference-guide.pdf Accessed 06/08/19

Iezzoni, L.I., 2009. 3. 1 Risk adjustment for performance measurement. *Performance measurement for health system improvement: experiences, challenges and prospects*, *251*.

NHS England 2014. Five Year Forward View. Available at: https://www.england.nhs.uk/wp-content/uploads/2014/10/5yfv-web.pdf Accessed 06/08/19

NHS England 2013. Patient reported outcome measures in England: Update to reporting and case-mix adjusting hip and knee procedure data. Available at: https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2013/10/proms-meth-prim-revis.pdf Accessed 06/08/19

NHS England Website 2019. Available at: https://www.nhs.uk/using-the-nhs/about-the-nhs/friends-and-family-test-fft/ Accessed 06/08/19

NHS England 2014 The Friends and Family Test Guide 2014. Available at: https://assets.nhs.uk/prod/documents/FFTGuide_Final_1807_FINAL.pdf Accessed 06/08/19

NHS England Website 2019. Patient Activation Measure (PAM). Available at: https://www.england.nhs.uk/personalisedcare/supported-self-management/patient-activation-measure/ Accessed 05/11/19

NHS England Website 2020. First Contact Physiotherapists. Available at: https://www.england.nhs.uk/gp/our-practice-teams/first-contact-physiotherapists/ Accessed 21/01/20

NICE 2019. Available at: https://www.nice.org.uk/Media/Default/Standards-and-indicators/indicators-consultation-2019.pdf Accessed 05/11/19

ONS 2019. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity Accessed 06/08/2019

Rolfson, O., Wissig, S., van Maasakkers, L., Stowell, C., Ackerman, I., Ayers, D., Barber, T., Benzakour, T., Bozic, K., Budhiparama, N. and Caillouette, J., 2016. Defining an international standard set of outcome measures for patients with hip or knee osteoarthritis: consensus of the international consortium for health outcomes measurement hip and knee osteoarthritis working group. *Arthritis care & research*, *68*(11), pp.1631-1639.

Appendix 5-3-1: Keele's Recommended List of Patient Experience Questions

Answer the following questions thinking about the consultation you have just had…

Item name	Question	Response options	Source of item
1. Treated with care and respect	The clinician listened attentively to what I said	Strongly agree 5 Agree 4 Uncertain 3 Disagree 2 Strongly disagree 1 Not applicable 9	Valuing patients as individuals
	The clinician was very approachable and easy to talk to	Strongly agree 5 Agree 4 Uncertain 3 Disagree 2 Strongly disagree 1 Not applicable 9	Valuing patients as individuals
	The clinician treated me kindly	Strongly agree 5 Agree 4 Uncertain 3 Disagree 2 Strongly disagree 1 Not applicable 9	Valuing patients as individuals

2. Being understood and valued	My problems were regarded as important	Strongly agree 5 Agree 4 Uncertain 3 Disagree 2 Strongly disagree 1 Not applicable 9	Valuing patients as individuals
	All of my questions were answered	Strongly agree 5 Agree 4 Uncertain 3 Disagree 2 Strongly disagree 1 Not applicable 9	Valuing patients as individuals
	I was treated as an intelligent human being	Strongly agree 5 Agree 4 Uncertain 3 Disagree 2 Strongly disagree 1 Not applicable 9	Valuing patients as individuals
3. Satisfaction	How likely are you to recommend this type of video consultation to friends and family if they need similar care or treatment?	Extremely likely 5 Likely 4 Neither likely nor unlikely 3 Unlikely 2 Extremely unlikely 1	Friends and family test

4. Confidence in clinical competence	How confident were you in the clinician's competency to assess and treat your problem?	Extremely 5 Very 4 Moderately 3 Slightly 2 Not at all 1	From the National FCP pilot
5. Shared decision-making	How much effort was made to help you understand your health issues?	Every effort was made 4 A lot of effort was made 3 Some effort was made 2 A little effort was made 1 No effort was made 0	From CollaboRATE
	How much effort was made to listen to the things that matter most to you about your health issues?	Every effort was made 4 A lot of effort was made 3 Some effort was made 2 A little effort was made 1 No effort was made 0	From CollaboRATE
	How much effort was made to include what matters most to you in choosing what to do next?	Every effort was made 4 A lot of effort was made 3 Some effort was made 2 A little effort was made 1 No effort was made 0	From CollaboRATE
6. Given sufficient information	Did you receive sufficient information about your condition or self-care?	Yes 1 No 0	From the National FCP pilot

7. Overall sense of improvement	Overall, how would you describe how you are compared to before the consultation?	Much better: 5 Better: 4 Same: 3 Worse: 2 Much worse: 1 Prefer not to say: 9	Global change item
8. Understanding of health condition	Thinking about your recent video consultation, how well do you feel it helped you understand your condition and any current treatment?	Completely: 4 Very well: 3 Moderately: 2 Slightly: 1 Not at all: 1	From MSK-HQ
9. Confidence to manage yourself	How confident do you now feel in being able to manage your health condition by yourself?	Extremely: 4 Very: 3 Moderately: 2 Slightly: 1 Not at all: 1	From MSK-HQ
10. Timeliness and convenience	How suitable was the timing of this video consultation for you?	Extremely: 4 Very: 3 Moderately: 2 Slightly: 1 Not at all: 1	

	How convenient was this video consultation for you?	Extremely 4 Very 3 Moderately 2 Slightly 1 Not at all 1	

Appendix 7-1: National PREM abstract (presented at the Centre for Advancing Practice conference by Paula Deacon).

Development of a Co-Produced Patient Reported Experience Measure for Community and Primary Care Musculoskeletal Services: A consensus Approach

Authors

Paula Deacon, Advance Physiotherapy Practitioner Lead, Community Musculoskeletal Assessment and Physiotherapy Service, The Dudley Group NHS Foundation Trust. Paula.deacon3@nhs.net

Dominic Ellington, MSK Operational Lead, Midlands Partnership NHS Foundation Trust Dominic.ellington@mpft.nhs.uk

Michael Brooks, Musculoskeletal Peer Support Worker, Midlands Partnership Foundation NHS Trust. Michael.Brooks@mpft.nhs.uk

Roanna Burgess, Consultant Physiotherapist/Musculoskeletal Service Lead, Sandwell and West Birmingham Hospitals NHS Trust. Doctoral Student, School of Medicine, Primary Care Centre Versus Arthritis, Keele University, roannaburgess@nhs.net

Dr Jonathan Hill, Reader in Physiotherapy, Director for Research for the School of Allied Health Professionals, School of Medicine, Primary Care Centre Versus Arthritis, Keele University j.hill@keele.ac.uk

Background

In the United Kingdom (UK), there is currently no standardised Musculoskeletal (MSK) Patient Reported Experience Measure (PREM) for use in community/primary care MSK services. MSK Services are therefore collecting different patient experience data leading to an inability to compare the experiences of MSK patients across services and settings. A standardised MSK PREM is important to ensure variation in the quality of care for MSK patients is minimised.

Aim

To gain consensus on a proposed set of domains and questions which would be used to develop a PREM for use in practice in community/primary care MSK services in the UK.

Method

A consensus process involving patients, healthcare professionals and researchers. Previous research and current MSK PREM questions used in these settings generated an initial list of proposed domains and questions. Proposed domains and questions were distributed for wider stakeholder consensus via online surveys.

Results

Sixty-six respondents completed the surveys (domain consensus and questions consensus). Out of thirteen domains, ten met consensus. Seven reached strong consensus, three moderate consensus and three didn't meet consensus so were excluded. Consensus domains were access/waiting times, shared decision making, confidence in clinicians, treated with care and respect, time and understanding, care planning, information giving, convenience/timeliness, needs met and ideas to improve our service.

Conclusion:

Ten PREM domains reached consensus. Domains were then populated with questions that also met either strong or moderate consensus. A co-produced MSK PREM has now been developed and will be piloted in a number of community/primary care settings across the UK.