Predictors of Functional Outcome in Musculoskeletal Healthcare: An Umbrella Review

Burgess, R. (1,2) Mansell, G. (1) Bishop, A. (1) Lewis, M. (1) Hill, J. (1)

1. Primary Care Centre Versus Arthritis, Research Institute for Primary Care & Health Sciences, Keele University, Keele, Staffordshire, ST5 5BG
2. Sandwell and West Birmingham Hospitals NHS Trust, Birmingham City Hospital, Dudley Road, Birmingham, B18 7QH

Corresponding Author: Roanna Burgess

Email: [roannaburgess@nhs.net](mailto:roannaburgess@nhs.net)

Tel: 01782 734707

Fax: 01782 734719

Category: Review Article

Funding Sources: This review received no funding. RB is supported by SWBH Your Trust Charityand a Keele ACORN studentship.

Conflict of Interest: The authors declare that there are no conflicts of interest.

**Significance**

This article presents an overview of systematic reviews and meta-analyses on predictors of functional outcome, measured using patient reported outcome measures (PROMs) in musculoskeletal healthcare. It highlights ‘generic’ patient factors that predict functional outcome across musculoskeletal conditions. Findings provide support for the development and content of generic musculoskeletal prognostic models including models used to case-mix adjust PROM data for baseline complexity. Generic musculoskeletal models and functional PROMs would facilitate more feasible comparison and benchmarking of musculoskeletal services.

# Abstract

Background: Multiple cohort and systematic review studies exist, reporting independent predictive factors associated with outcome in musculoskeletal populations. These studies have found evidence for a number of ‘generic’ factors that have been shown to predict outcome across musculoskeletal patient cohorts. This review provides a higher level review of the evidence with a focus on generic patient factors associated with functional musculoskeletal 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 musculoskeletal 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 musculoskeletal prognostic models, including models used for case-mix-adjustment of patient reported outcome measure data.

# Key Words

Predictors, musculoskeletal, functional outcome

# 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 (The Kings Fund, 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 musculoskeletal (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). A recent review of case-mix adjustment models (adjusting functional PROM scores) 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. 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.

## Searches

Medline, AMED, EMBASE and CINAHL electronic databases were searched alongside the Cochrane Database of systematic reviews (search strategy for Medline included within Appendix S1). 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.

## 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 2 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. 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 1: Level of evidence for generic prognostic factors for poor functional outcome (Adapted from Sacket (2000), Kooijman et al (2015), Artus et al (2017))

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 1 for flowchart of search results.

## Figure 1: Flowchart of Search Results

## Quality Assessment

The quality checklist is shown with results for included studies in Table 2. For each criterion within the checklist, studies were rated as met, not met, or not applicable (NA).

## 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 2). 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 1, 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 4 and 5. Where possible only functional outcome was reported unless this could not be separated from other primary outcomes within studies.

### Table 2: Quality of Included Studies (Critical appraisal checklist for systematic reviews and research syntheses (Aromataris et al, 2014))

### Table 3: Characteristics of Included Studies

### T**able 4: Studies showing evidence for predictive factors associated with functional outcome**.

### Table 5: Generic Predictors of Functional Outcome

## 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 4**.** 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 5. Similar predictors are grouped together where appropriate. Table 5 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 5 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.

# 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 5) 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) 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, 2018). 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**

Our 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).

Our 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 our 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, 2018), 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.

# Author Contributions

All persons who meet authorship criteria are listed as authors, and all authors certify that they have participated sufficiently in the work to take public responsibility for the content, including participation in the concept, design, analysis, writing, or revision of the manuscript.

# References

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. <http://joannabriggs.org/assets/docs/sumari/reviewersmanual-2014.pdf>

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*.

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>

Campbell, P., Hill, J. C., Protheroe, J., Afolabi, E. K., Lewis, M., Beardmore, R., ... & 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*, 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, (2010). PROMs risk adjustment methodology guide for general surgery and orthopaedic procedures. Northgate. Available:<https://www.england.nhs.uk/statistics/wpcontent/uploads/sites/2/2013/07/proms-ris-adj-meth-sur-orth.pdf>

Deeks, J. J., Dinnes, J., D'amico, R., Sowden, A. J., Sakarovitch, C., Song, F., ... & Altman, D. G. (2003). Evaluating non-randomised intervention studies. *Health technology assessment (Winchester, England)*, *7*(27), iii-x.

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.

Guyatt, G., Oxman, A. D., Akl, E. A., Kunz, R., Vist, G., Brozek, J., ... & Rind, D. (2011). GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. *Journal of clinical epidemiology*, *64*(4), 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 18-P-93066/9-01. Health & Human Services/Centers for Medicare & Medicaid Services. Available: http:// [www.cms.hhs.gov/TherapyServices/downloads/P4PFinalReport06- 01-06.pdf](http://www.cms.hhs.gov/TherapyServices/downloads/P4PFinalReport06-%2001-06.pdf)

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., ... & 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), 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.

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. (2012). PROMs in England; the case-mix adjustment methodology. <https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/216507/dh_133449.pdf>

NHS England. (2013). Patent Reported Outcome Measures (PROMs). An alternative aggregation methodology for case-mix adjustment. <http://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2013/07/proms-agg-meth-adju.pdf>

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., ... & Ferreira, P. H. (2016). Symptoms of depression as a prognostic factor for low back pain: a systematic review. *The Spine Journal*, *16*(1), 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.

Riley, R. D., Hayden, J. A., Steyerberg, E. W., Moons, K. G., Abrams, K., Kyzas, P. A., ... & Hemingway, H. (2013). Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. *PLoS medicine*, *10*(2), e1001380.

Sackett, D.L., (2000). Therapy. *Evidence based medicine. How to practice and teach EBM*, 136-137.

Shea, B. J., Grimshaw, J. M., Wells, G. A., Boers, M., Andersson, N., Hamel, C., ... & Bouter, L. M. (2007). Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC medical research methodology*, *7*(1), 10.

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.

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.