ELSEVIER

Physiotherapy

Physiotherapy xxx (2018) xxx–xxx

Systematic review

Models used for case-mix adjustment of patient reported outcome measures (PROMs) in musculoskeletal healthcare: A systematic review of the literature

R. Burgess [a,b,*], A. Bishop [a], M. Lewis [a], J. Hill [a]

[a] *Arthritis Research UK Primary Care Centre, Research Institute for Primary Care & Health Sciences, Keele University, Staffordshire, ST5 5BG, United Kingdom*

[b] *Sandwell and West Birmingham Hospitals NHS Trust, Dudley Road, Birmingham, B18 7QH, United Kingdom*

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 musculoskeletal settings.

Keywords: case-mix adjustment model; musculoskeletal; patient outcomes; PROM

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 [1]. For example, the UK National PROMs Programme has successfully raised standards in the area of hip and knee replacement surgery [2]. Patient outcomes are a function of; therapeutic intervention effectiveness, quality of care, patient attributes that affect their response to care (e.g. 'risk factors'), the natural course of a condition and random chance [3,4]. 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 [3]. For

* Corresponding author at: Arthritis Research UK Primary Care Centre, Research Institute for Primary Care & Health Sciences, Keele University, Staffordshire, ST5 5BG, United Kingdom.

E-mail address: r.m.burgess@keele.ac.uk (R. Burgess).

https://doi.org/10.1016/j.physio.2018.10.002

0031-9406/

example to enable fair comparisons across different musculoskeletal physiotherapy services it may be appropriate to adjust for population differences in age or symptom duration, as these are known to influence patient outcomes following treatment [5]. Other known patient factors that influence musculoskeletal treatment outcomes include; gender, symptom severity, and impairment type [6]. These patient factors are beyond the control of the treatment provider, unlike provider factors such as the waiting time, clinic setting, or treating clinician, which also influence treatment outcomes [7]. 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 [8]. For example, if one physiotherapy service had treating clinicians of a much higher grade than another, and grade of therapist was adjusted for when examining their respective treatment outcomes, then any variation due to the differing skill-mix between the services would be adjusted out rather than being used to help explain the differences and inform quality improvement initiatives. Most case-mix models therefore only adjust for patient factors to allow for fair inter provider comparisons [8].

Within a musculoskeletal context the evidence for case-mix adjustment models to compare inter provider treatment outcomes 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 musculoskeletal 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 [9], and has been registered on the PROSPERO database (CRD42017055948).

Eligibility Criteria

Inclusion criteria were: studies from January 1992 to May 2017 (in line with early implementers of musculoskeletal PROM collection [10] and to provide currency and applicability of results), English language studies (due to resource limits), observational cohort studies, adult patients seeking treatment for musculoskeletal conditions, use of a case-mix adjustment model (focus on development, refinement or validation), self-reported treatment outcomes at a follow-up time-point (capturing treatment effect/change), and models adjusting PROMs and including variables feasible for widespread collection (not using variables such as imaging results that are not uniformly collected). Exclusions were: studies not reporting detailed results, and those not reporting statistical model effectiveness.

Searches

A search strategy was developed iteratively with guidance from an experienced systematic reviewer, initially conducting test searches for a single database until the refined strategy was agreed that amalgamated sets of search terms, reduced individual terms, and exploded terms such as 'musculoskeletal' to optimise the balance between search sensitivity and precision [11]. Search-terms included key words for; target population; musculoskeletal conditions; outcomes; and methodology. Electronic databases searched were: CINAHL; MEDLINE; EMBASE; AMED and HMIC (see Appendix A for search strategy (MEDLINE)) from January 1992 to May 2017. Grey literature included searches of NHS Evidence websites of the Department of Health [12] and NICE [13]. Additional searches included references and citations of included studies. Seminal authors/research groups were also contacted for all identified case-mix models to ensure latest iterations were included and to identify any additional models.

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 confirm they met the inclusion criteria.

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 [14], and quality assessed using the CASP Cohort Quality Tool [15]. Agreement on study inclusion was first discussed between two reviewers. As there were no disputed studies discussion for agreement between all reviewers was not required.

Data synthesis

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. A meta-analysis pooling the study data was not possible due to the large methodological diversity (heterogeneity) among studies [11] in patient factors and statistical methods used. For this reason, results were summarised in tables and discussed in detail. Each case-mix adjustment model and their associated studies/papers and statistical methods were presented together for ease of viewing overarching findings.

Please cite this article in press as: Burgess R, et al. Models used for case-mix adjustment of patient reported outcome measures (PROMs) in musculoskeletal healthcare: A systematic review of the literature. Physiotherapy (2018), https://doi.org/10.1016/j.physio.2018.10.002

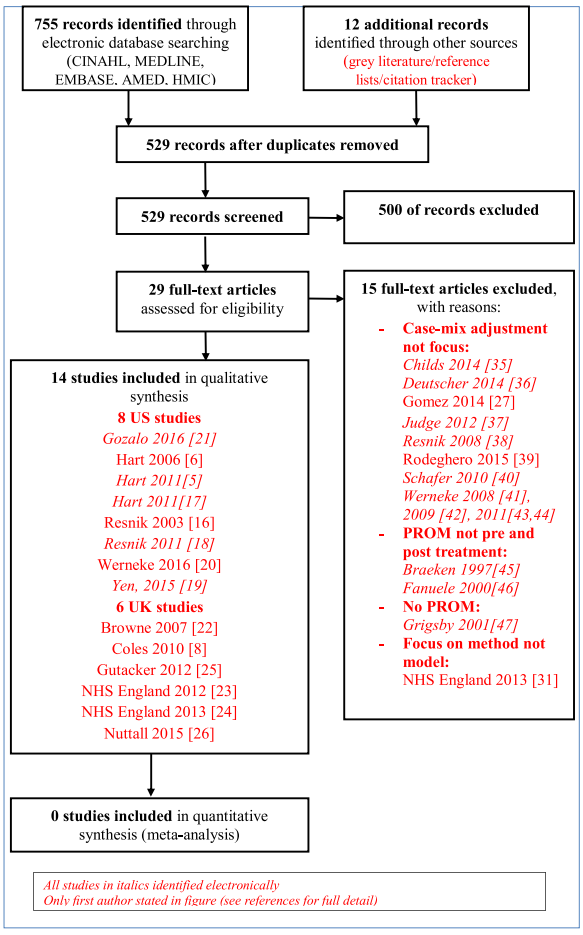

Fig. 1. Flowchart of Search Results [35–47].

Results

Search Results

Electronic database searches identified 755 articles for consideration with 517 remaining after duplicate removal (see Fig. 1). Grey literature and additional searching identified a further 12 articles. All seven experts (or alternative experts from their research group) responded and this identified one additional manuscript that was being prepared for submission that was unable to be included within the review. Following screening, fourteen articles were included (see Fig. 1). Two broad case-mix adjustment models were identified; US Focus on Therapeutic Outcomes (FOTO), and UK National PROMs (NPROMs).

Eight of the fourteen studies included were undertaken in the US, using data from the FOTO database [5,6,16–19], with four of those authored (primary author) by members of the FOTO Research Advisory Board (FRAB) [5,6,17,20]. The other four were independently led and given access to the FOTO database [16,18,19,21] although two of them were also co-authored by FRAB members [16,18]. Included study sample sizes ranged from n = 323 [17] to n = 189,088 [5]. The pooled sample size across US studies with pre and post treatment data was 546,726.

Six of the fourteen included studies were UK based. These included feasibility work for the NHS England NPROMs Programme [22], NPROMs publications [8,23,24], and independent researchers using NPROMs data [25,26]. All of these studies were only identified following review of the grey literature/additional searches as they were all NHS publications or secondary analyses of NHS data. Included study sample sizes ranged from n = 387 [22] to NPROMs data which increased yearly from; 2009-10 (n = 85,177), 2010-11 (n = 95,406), 2011-12 (n = 101,454) totalling 282,037 patients [23,24]. 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 [8,18,22], and the rest undertaking retrospective analyses of existing datasets. For results detail see Table 1 for quality of included studies, Table 2 for summary of articles included, and Table 3 for summary of model variables within included studies.

Quality Appraisal

The CASP quality evaluation found that studies were of a good quality (see Table 1). There were however consistent sources of bias across 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 a large percentage of participants with missing data (see Table 1). Hart et al. [17] 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 biased their sample to those more likely to be psychologically impaired (as acknowledged by the authors). Three of the eight US studies did however use inverse probability weighting to account for missing data [18,19,21]. 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 these were unlikely to lead to systematic selection biases as missing data included both patients with characteristics associated with better and worse outcome [5,6,17,20]. 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 [10]). 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

Please cite this article in press as: Burgess R, et al. Models used for case-mix adjustment of patient reported outcome measures (PROMs) in musculoskeletal healthcare: A systematic review of the literature. Physiotherapy (2018), https://doi.org/10.1016/j.physio.2018.10.002

Table 1
Quality Assessment Using CASP Cohort Tool.

CASP Cohort Tool													
Author	Clearly focussed	Recruitment acceptable	Exposure accurately measured	Outcome accurately measured	Identified confounding	Accounted for confounding	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 [16]	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	
Hart 2006 [6]	Yes	CT	Yes	Yes	Yes.	Yes	Yes	Yes	Yes	CT	No	CT	62
Hart 2011 [5]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	62
Hart 2011 [17]	Yes	No	Yes	Yes	Yes	Yes	No	CT	No	Yes	Yes	CT	0.80
Resnik 2011 [18]	Yes	No	Yes	Yes	Yes	Yes	No	CT	Yes	Yes	Yes	Yes	44.30
Yen 2015 [19]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Gozalo 2016 [21]	Yes	Yes	Yes	Yes	Yes	Yes	No	CT	Yes	Yes	Yes	Yes	57.20
Werneke 2016 [20]	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes	35
First Author UK													
Browne 2007 [22]	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	CT	90.2-91.6
Coles 2010 [8]	Yes	CT	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
DoH 2012 [23]	Yes	CT	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
DoH 2013 [24]	Yes	CT	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Gutacker 2012 [25]	Yes	CT	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	
Nuttall 2015 [26]	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	83.90

FU; follow up, CT; Can't tell.

Please cite this article in press as: Burgess R, et al. Models used for case-mix adjustment of patient reported outcome measures (PROMs) in musculoskeletal healthcare: A systematic review of the literature. Physiotherapy (2018), https://doi.org/10.1016/j.physio.2018.10.002

Table 2
Summary of Included Studies.

First Author/s	Design	Setting	Data Sources	Study Size (complete/ included datasets)	PROMs	Number of variables	Model R2 (where available)
US Studies							
Resnik and Hart (2003) [16]	Retrospective cohort	Outpatient physical therapy	FOTO	24,276	OHS, SF-12, SF-36	8	35-42%
Hart and Connolly (2006) [6]	Retrospective cohort	Outpatient therapy	FOTO	189,088	FS	12	35-36%
Hart (2011) [5]	Retrospective cohort	Outpatient therapy	FOTO	49,376	FS	8	30%
Hart (2011) [17]	Prospective cohort	Outpatient therapy	FOTO	257	FS	10 (plus PM)	31% (intake model)
Resnik (2011) [18]	Prospective cohort	Outpatient therapy	FOTO	44,925	FS	8	18-40%
Yen (2015) [19]	Retrospective cohort	Outpatient therapy	FOTO	147,623	FS	7	31% (FE model)
Werneke (2016) [20]	Retrospective cohort	Outpatient physical therapy	FOTO	723	FS	13 (tested in BM)	35% (BM)
Gozalo (2016) [21]	Retrospective cohort	Outpatient therapy	FOTO	90,392	FS	8	
UK Studies							
Browne (2007) [22]	Prospective cohort	Orthopaedic		700	EQ5D Index, OHS, OKS, SF-36	8	24-27%
Coles (2010) [8]	Prospective cohort	Orthopaedic	NPROMs	29759	EQ5D Index, EQ5D VAS, OHS, OKS	16-20 dependent on PROM model	23-30%
NHS England (2012) [23]	Retrospective cohort	Orthopaedic	NPROMs	282,037	EQ5D Index, EQ5D VAS, OHS, OKS	13-15 dependent on tool (some variable items listed & coded separately)	
Gutacker (2012) [25]	Retrospective cohort	Orthopaedic	NPROMs	24,568	EQ5D	7	
NHS England (2013) [24]	Retrospective cohort	Orthopaedic	NPROMs	282,037 (as for NHS England, 2012)	EQ5D Index, EQ5D VAS, OHS, OKS	12 (some variable items listed & coded separately)	
Nuttall (2015) [26]	Retrospective cohort	Orthopaedic	NPROMs	30,555	OKS	10 (some variable items listed & coded separately)	26% (OLS and FE model)

OHSM; Overall health status measure, SF-12; Short Form 12, SF-36; Short Form 36, FS; Functional Status, PM; psychological measure, BM; baseline model, OLS; ordinary least squares, FE; fixed effects, OKS; Oxford Knee Score, OHS; Oxford Hip Score.

quantify estimates of efficacy for a given time. Patients with missing follow up data may therefore be 'missing not at random' [27] having chosen to cease attending leading to further potential attrition bias [11]. Resnik and Hart [16] reported that these patients were younger and had higher functional status scores and therefore hypothesised that they were likely to have ceased attending due to resolution of their symptoms. However, not including those with greater chances of improvement as well as the variation in outcome collection timing could substantially impact on the case-mix models and their reported predictive abilities [28].

Key sources of bias across UK studies included: Selection-bias due to the exclusion of those with missing data (see Table 1). The study by Browne et al. [22] used the SF-36 rule [29] for dealing with missing data, but 25% of eligible patients were excluded due to failure to invite these 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 have potentially biased 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 [30]. Whether this impacted on results would depend on whether unlinked or missing data was missing at random [27] or whether this was due to systematic poor administrative processes at certain

Please cite this article in press as: Burgess R, et al. Models used for case-mix adjustment of patient reported outcome measures (PROMs) in musculoskeletal healthcare: A systematic review of the literature. Physiotherapy (2018), https://doi.org/10.1016/j.physio.2018.10.002

Table 3
Summary of Risk-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 [18]	x*	x	x		x*	x	x*			x						
Hart and Connolly 2006 [6]	x*	x*	x		x*	x	x	x		x						x
Hart 2011 [5]	x*	x	x	x*	x*	x	x								x	
Hart 2011 [17]	x*	x	x	x*	x	x*	x*			x					x	x
Resnik 2011 [18]	x	x	x	x	x		x			x						
Yen 2015 [19]	x	x	x	x	x	x	x									
Gozalo 2016 [21]	x*	x	x	x*	x*	x	x*	x							x	
Werneke 2016 [20]	x*	x*	x	x*	x*	x*	x*			x						x
First Author/s UK																
Browne 2007 [22]	x*	x*	x	x*	x	x*			x*							
Coles 2010 [8]	x*	x	x	x*	x	x		x	x*		x	x	x*	x		
NHS England 2012 [23]	x*	x	x	x	x	x		x	x*		x	x*	x*	x		
NHS England 2013 [24]	x*	x*	x*	x*	x			x*	x*		x*	x*	x*	x		
Gutacker 2012 [25]	x	x	x*	x*		x*		x*	x*							
Nuttall 2015 [26]	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 (2011) * 3 largest predictors.
Hart et al (2011) * 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).
NHS England (2012) * 4 most predictive across models (not including depression).
NHS England (2013) * 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).

Please cite this article in press as: Burgess R, et al. Models used for case-mix adjustment of patient reported outcome measures (PROMs) in musculoskeletal healthcare: A systematic review of the literature. Physiotherapy (2018), https://doi.org/10.1016/j.physio.2018.10.002

provider NHS trusts, which is unclear. Follow-up data collection across UK studies was standardised at a six month time-point although baseline data collection occurred both at pre admission clinics and at admission for the surgical procedure, leading to a small source of variation.

All included studies used data from clinical databases and were therefore impacted by limitations in controlling the quality of the data and rates of attrition. Most studies reported these limitations reinforcing the issues around the use of clinical data for research purposes. However although acknowledged, these limitations led to a high risk of bias for this domain within included studies [11].

Model development history

US Model

Early FOTO models made case-mix adjustments using 12 baseline variables as demonstrated by Hart and Connolly [6] (see Table 3) that were found to have a significant effect on discharge functional status (FS). This model predicted 35% of total variance, meaning that 35% of the variance in post treatment outcome could be explained by the model. The three most important patient factors in their model were; baseline FS, age and symptom duration [6], supporting work from Resnik and Hart [16]. FOTO Inc. later moved to a case-mix adjustment model with eight patient factors, aware of the need to balance model performance with data collection feasibility [6], as demonstrated in the paper by Hart et al. [5], who looked at the benefit of adding fear avoidance beliefs (FABQ-PA) to the model. Their results demonstrated R2 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 not recommending this variable for model inclusion.

UK Model

In 2007 Browne et al. [22], set out to determine the feasibility of collecting pre and post-operative outcome data from patients undergoing elective surgery, and to develop methods to analyse and present the pooled data from different hospitals. Elective surgeries included five areas, with two of musculoskeletal interest: unilateral hip replacement and unilateral knee replacement. Significant variables within their case-mix adjustment models were baseline PROM score, comorbidities, general health, surgical history, age, and Index of Multiple Deprivation (IMD). Models explained between 24% and 27% of total variance in treatment outcome.

Following the feasibility work by Browne et al. [22], Coles [8] published the full UK NPROMs case-mix adjustment methodology (see Table 3 for list of variables). Coles [8] describes six orthopaedic models (separate models for each PROM used and for each intervention). Models ranged from 16-20 included variables and explained between 23% and 30% of total variance. All models found the patient's baseline score to be highly predictive of outcome, as well as IMD, comorbidities, patients reporting themselves free of a disability (positive impact), and general health.

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 general health [23]. Key predictive variables within the updated model were baseline PROM score, disability status, comorbidity of depression, patient needing assistance with questionnaire, and IMD [23]. In 2013, an alternative aggregation model (AAM) was proposed by NHS England [31], 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 model changes included removing the previous surgery variable and inclusion of some 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 [24].

Model validation

US Model

Hart and Connolly [6] 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 development sample the predicted discharge FS was very close to the actual discharge FS (average predictive ratio 1.045), although the model slightly over predicted FS in the second testing sample. The paper by Hart et al. [5] 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 [5].

UK Model

The inception NPROMs paper [8] 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 model for Knee surgery using EQ5D VAS as the outcome showed the only significant difference in samples. This was due to the low incidence of some comorbidities, and lack of specific admission and discharge data. All models showed face validity containing appropriate variables with directionally expected coefficients. Nuttall et al. [26] independently reviewed case-mix adjustment of NPROMs data. Mean predicted post-operative scores and mean actual scores were compared using three statistical methods (ordinary least squares (OLS), fixed effects (FE), and random

Please cite this article in press as: Burgess R, et al. Models used for case-mix adjustment of patient reported outcome measures (PROMs) in musculoskeletal healthcare: A systematic review of the literature. Physiotherapy (2018), https://doi.org/10.1016/j.physio.2018.10.002

effects (RE) models). They demonstrated that a fixed effects (FE) model performed the best [26].

Model statistical methods

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 [6], 0.1 [20] and 0.15 [8]) 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 (R2) [6,8]. Hierarchical models were demonstrated in later papers [19,21,27]. UK NPROMs moved to the use of a generalised least squares (GLS) method in 2011 [23]. Support is growing for the use of GLS [23,26] and hierarchical mixed models [19] 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 R2 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) [32]. Predictive ability across US study models ranged from 18-42% [5,6,16,17,18,19,20] and in UK models from 23-30% [8,22,26], demonstrating moderate to strong predictive ability across models [33].

Discussion

Table 3 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 widely used variables across models predicting outcome 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 musculoskeletal case-mix adjustment modelling. Variables such as exercise history, living alone, FABQ, use of medication, and pain intensity had some limited support but 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 [19,34].

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 outcome tools and different timing of follow-up data 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 and treatment outcomes are measured can have significant effects on model predictive performance [28].

Limitations of the review

The review focussed on the case-mix adjustment of musculoskeletal PROM data. However, the outcome used within studies was not limited and therefore studies and the predictive performance of models identified cannot be fully compared. Evidence from the UK NPROMs research demonstrates that different variables are necessary dependent on the outcome used [8,24]. 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, treatments and outcomes across settings all vary significantly. The review was also limited to English language publications meaning that there may be models reported in languages other than English that have not been included.

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 these existing models within and across musculoskeletal settings and countries, in order to be able to implement these models across healthcare settings.

Recommendations for future case-mix adjustment modelling of musculoskeletal PROMs based on the combined study findings 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 [6,8,20].
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.

Please cite this article in press as: Burgess R, et al. Models used for case-mix adjustment of patient reported outcome measures (PROMs) in musculoskeletal healthcare: A systematic review of the literature. Physiotherapy (2018), https://doi.org/10.1016/j.physio.2018.10.002

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 musculoskeletal 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 musculoskeletal health outcomes. This will allow for effective performance profiling and future benchmarking of musculoskeletal services, both nationally and internationally.

Contribution of the Paper

- This systematic review has identified two broad musculoskeletal 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 musculoskeletal clinical pathways of care will allow for further development of performance profiling and benchmarking across musculoskeletal practice, with the aim of improving quality and equity of musculoskeletal healthcare provision.

Ethical Approval: Not applicable.

Funding: Not applicable.

Conflict of Interest: This is to confirm that co-author Annette Bishop is an Editor for Physiotherapy but was not involved with the peer review of the paper or the final decision.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.physio.2018.10.002.

References

[1] Black N. Patient reported outcome measures could help transform healthcare. BMJ: British Medical Journal (Online) 2013;346.

[2] NHS England Bite-size guide to patient insight: The National Patient Reported Outcome Measures PROMS Programme. 2016. Available from: https://www.england.nhs.uk/wp-content/uploads/2018/08/proms-guide-aug-18-v3.pdf.

[3] Iezzoni 3 LI. 1 Risk adjustment for performance measurement. Performance Measurement for Health System Improvement: Experiences. Challenges and Prospects 2009;251.

[4] Vasseljen O, Woodhouse A, Bjørngaard JH, Leivseth L. Natural course of acute neck and low back pain in the general population: the HUNT study. PAIN® 2013;154(Aug (8)):1237–44.

[5] Hart DL, Werneke MW, Deutscher D, George SZ, Stratford PW. 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 2011;41(May (5)):336–45.

[6] Hart DL, Connolly JB. Pay-for-performance for physical therapy and occupational therapy: Medicare Part B Services. Final report. Grant. 2006 Jun 1:9-01.

[7] Werneke M, Hart DL. Centralization phenomenon as a prognostic factor for chronic low back pain and disability. Spine 2001;26(Apr (7)):758–64.

[8] Coles J. PROMs risk adjustment methodology guide for general surgery and orthopaedic procedures. Northgate Information Solutions (UK) Ltd. 2010. Available from: https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2013/07/proms-ris-adj-meth-sur-orth.pdf.

[9] Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Systematic reviews 2015;4(Jan (1)):1.

[10] Focus on Therapeutic Outcomes Inc. 2018 [cited August 2018] Available from: https://www.fotoinc.com/about-foto.

[11] Higgins JPT, Green S. (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0:130 [updated March 2011]. The Cochrane Collaboration, 2011. Available from: www.handbook.cochrane.org.

[12] Department of Health [Accessed May 2017] Available from: https://www.gov.uk/government/publications?departments%5B%5D=department-of-health.

[13] NICE [Accessed May 2017] Available from: www.evidence.nhs.uk.

[14] Von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. PLoS Med 2007;4:e296.

[15] Critical Appraisal Skills Programme. CASP (Cohort Study) Checklist. 2017. [online] Available from: http://docs.wixstatic.com/ugd/dded87_5ad0ece77a3f4fc9bcd3665a7d1fa91f.pdf.

[16] Resnik L, Hart DL. Using clinical outcomes to identify expert physical therapists. Physical Therapy 2003;83(Nov (11)):990–1002.

[17] Hart DL, Werneke MW, Deutscher D, George SZ, Stratford PW, Mioduski JE. Using intake and change in multiple psychosocial measures to predict functional status outcomes in people with lumbar spine syndromes: a preliminary analysis. Physical therapy 2011;91(Dec (12)):1812–25.

[18] Resnik L, Gozalo P, Hart DL. Weighted index explained more variance in physical function than an additively scored functional comorbidity scale. Journal of clinical epidemiology 2011;64(Mar (31)):320–30.

[19] Yen SC, Corkery MB, Chui KK, Manjourides J, Wang YC, Resnik LJ. Risk adjustment for lumbar dysfunction: comparison of linear mixed models with and without inclusion of between-clinic variation as a random effect. Physical therapy 2015;95(Dec (12)):1692–702.

[20] Werneke MW, Edmond S, Deutscher D, Ward J, Grigsby D, Young M, McGill T, McClenahan B, Weinberg J, Davidow AL. 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 2016;46(Sep (9)):726–41.

Please cite this article in press as: Burgess R, et al. Models used for case-mix adjustment of patient reported outcome measures (PROMs) in musculoskeletal healthcare: A systematic review of the literature. Physiotherapy (2018), https://doi.org/10.1016/j.physio.2018.10.002

[21] Gozalo PL, Resnik LJ, Silver B. Benchmarking outpatient rehabilitation clinics using functional status outcomes. Health services research 2016;51(2):768–89.
[22] Browne J, Jamieson L, Lewsey J, van der Meulen J, Black N, Cairns J, Lamping D, Smith S, Copley L, Horrocks J. 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; 2007.
[23] NHS England. PROMs in England; the case-mix adjustment methodology. 2012. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/216507/dh_133449.pdf.
[24] NHS England. Patient reported outcome measures in England: Update to reporting and case-mix adjusting hip and knee procedure data. 2013. Available from: https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2013/10/proms-meth-prim-revis.pdf.
[25] Gutacker N, Bojke C, Daidone S, Devlin N, Street A. Analysing hospital variation in health outcome at the level of EQ-5D dimensions. CHE Research Paper, no. 74. York, UK: Centre for Health Economics, University of York; 2012.
[26] Nuttall D, Parkin D, Devlin N. Inter-provider comparison of patient-reported outcomes: developing an adjustment to account for differences in patient case mix. Health economics 2015;24(1):41–54.
[27] Gomes M, Gutacker N, Bojke C, Street A. Addressing missing data in patient-reported outcome measures (PROMs): implications for comparing provider performance (No. 101cherp). Centre for Health Economics, University of York; 2014.
[28] Whittle R, Royle KL, Jordan KP, Riley RD, Mallen CD, Peat G. Prognosis research ideally should measure time-varying predictors at their intended moment of use. Diagnostic and Prognostic Research 2017;1(Dec (1)):1.
[29] Ware John, Snow Kk, Kosinski MA, Gandek BG. SF36 Health Survey: Manual and Interpretation Guide. Lincoln, RI: Quality Metric, Inc; 1993. p. 30.
[30] HSCIC. Finalised Patient Reported Outcome Measures (PROMs) in England: April 2012 to March 2013. 2014. Available from: https://www.gov.uk/government/statistics/patient-reported-outcome-measures-proms-in-england-finalised-april-2012-to-march-2013.
[31] NHS England Patent Reported Outcome Measures (PROMs) An alternative aggregation methodology for case-mix adjustment. 2013. Available from: http://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2013/07/proms-agg-meth-adju.pdf.
[32] Petrie A, Sabin C. Medical statistics at a glance. John Wiley & Sons; 2013. Nov 8. 84.
[33] Cohen J. Statistical power analysis for the behavioral sciences. 2nd. 79-82.
[34] Burstin HR, Lipsitz SR, Brennan TA. Socioeconomic status and risk for substandard medical care. Jama 1992;268(Nov (17)):2383–7.
[35] Childs JD, Harman JS, Rodeghero JR, Horn M, George SZ. Implications of practice setting on clinical outcomes and efficiency of care in the delivery of physical therapy services. journal of orthopaedic & sports physical therapy 2014;44(Dec (12)):955–63.
[36] Deutscher D, Werneke MW, Gottlieb D, Fritz JM, Resnik L. Physical therapists' level of McKenzie education, functional outcomes, and utilization in patients with low back pain. journal of orthopaedic & sports physical therapy 2014;44(Dec (12)):925–36.
[37] Judge A, Javaid MK, Arden NK, Cushnaghan J, Reading I, Croft P, Dieppe PA, Cooper C. Clinical tool to identify patients who are most likely to achieve long-term improvement in physical function after total hip arthroplasty. Arthritis care & research 2012;64(6):881–9.
[38] Resnik L, Liu D, Mor V, Hart DL. Predictors of physical therapy clinic performance in the treatment of patients with low back pain syndromes. Physical therapy 2008;88(Sep (9)):989–1004.
[39] Rodeghero JR, Cook CE, Cleland JA, Mintken PE. Risk stratification of patients with low back pain seen in physical therapy practice. Manual therapy 2015;20(Dec (6)):855–60.
[40] Schäfer T, Krummenauer F, Mettelsiefen J, Kirschner S, Günther KP. Social, educational, and occupational predictors of total hip replacement outcome. Osteoarthritis and cartilage 2010;18(Aug (8)):1036–42.
[41] Werneke MW, Hart DL, Resnik L, Stratford PW, Reyes A. Centralization: prevalence and effect on treatment outcomes using a standardized operational definition and measurement method. journal of orthopaedic & sports physical therapy 2008;38(Mar (3)):116–25.
[42] Werneke MW, Hart DL, George SZ, Stratford PW, Matheson JW, Reyes A. Clinical outcomes for patients classified by fear-avoidance beliefs and centralization phenomenon. Archives of Physical Medicine and Rehabilitation 2009;90(May (5)):768–77.
[43] Werneke MW, Hart DL, George SZ, Deutscher D, Stratford PW. 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 2011;41(Dec (12)):969–80.
[44] Werneke MW, Hart DL, Cutrone G, Oliver D, McGill MT, Weinberg J, Grigsby D, Oswald W, Ward J. Association between directional preference and centralization in patients with low back pain. journal of orthopaedic & sports physical therapy 2011;41(Jan (1)):22–31.
[45] Braeken AM, Lochhaas-Gerlach JA, Gollish JD, MYĹES JD, Mackenzie TA. Determinants of 6—12 Month Postoperative Functional Status and Pain After Elective Total Hip Replacement. International Journal for Quality in Health Care 1997;9(Jan (6)):413–8.
[46] Fanuele JC, Birkmeyer NJ, Abdu WA, Tosteson TD, Weinstein JN. The impact of spinal problems on the health status of patients: have we underestimated the effect? Spine 2000;25(Jun (12)):1509–14.
[47] Grigsby J, Kaye K, Kowalsky J, Kramer AM. Relationship between functional status and the capacity to regulate behavior among elderly persons following hip fracture. Rehabilitation Psychology 2002;47(Aug (3)):291.

Please cite this article in press as: Burgess R, et al. Models used for case-mix adjustment of patient reported outcome measures (PROMs) in musculoskeletal healthcare: A systematic review of the literature. Physiotherapy (2018), https://doi.org/10.1016/j.physio.2018.10.002