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

### 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 some 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 musculoskeletal (MSK) treatment outcome (Burgess et al, 2020). 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.

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 positive or negative ‘outliers’. For example; specific arthroplasty implant brands, early post-operative mobilisation regimes, and enhanced wound management protocols have been shown to deliver better outcomes for patients undergoing joint replacement surgery. Significant improvements in clinical outcomes were realised across providers with the implementation of these 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) 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) and is called the National PROMs (NPROMs) model within this paper, 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 all 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 paper.

This study therefore 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.

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).

2; Develop a new case-mix adjustment model (Keele Model 1) using an evidence-synthesis approach (informed by a recent systematic review (Burgess et al, 2019) and umbrella review (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 study 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 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) and Tampa Scale for Kinesiophobia (TSK) (FAB-TSK PROM), see **Table 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 regression analysis (see equation).

Multiple linear regression equation with k predictor variables: y = β0 + β1X1 + β2X2 + ··· βkXk + E

**Table 1:** **Variables in STarT MSK Trial Dataset & those used in NPROMs and FOTO Models.**

**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 within the modified FOTO model were: age (continuous), sex (binary), baseline functional status (FS) (MSK-HQ score (continuous) replaced FOTO FS computerised adaptive testing (CAT) PROM ), pain site (5 categories (replaced FOTO body part and care type i.e., orthopaedic) ), duration of symptoms (7 categories (how long since whole month without pain) replaced 6 FOTO acuity categories categorized as number of days from onset of the treated condition), comorbidities (4 categories), previous pain episodes (5 categories (replaced FOTO 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 rehabilitation 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 power was then calculated and displayed showing R2 and adjusted R2 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). The EQ5D5L is a standardised measure of health status (EuroQol, 2019). This PROM was used to additionally evaluate both of the existing models (modified FOTO/NPROMs) due to the MSK-HQ PROM being different to PROMs used in these internally validated models.

**Modified NPROMs Model**

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

Available variables in the STarT MSK Trial dataset included within the modified NPROMs model were: 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 categories (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 (Coles, 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, NHSE 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 power was then calculated and displayed showing changes in R2 and F values. The modified model power was also compared with a model using an alternative PROM (EQ5D5L) to evaluate if R2 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).

Independent variables were force entered in batches based on evidence-based recommendations (see **Table 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 2: Recommendations for variables to include in MSK case-mix adjustment model development**

**Keele Model 2: Statistically informed model**

The full list of available variables within the STarT MSK Trial dataset listed above (see Table 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 R2 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 we repeated regression analyses using mixed random effects (RE) models using STATA statistical software to identify if this impacted on coefficient values/significance.

### Results:

Descriptive data for the STarT MSK Trial dataset is presented in **Table 3** providing mean or percentage values for each variable that was entered into multivariate models, standard deviation (SD), and number of participants (n) alongside % of participants with available data for each variable. **Table 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 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 3**), this may explain why this variable was not predictive in this cohort.

**Table 3: Descriptive Statistics**

**Table 4: Univariate Analysis**

**Objective 1:**

**Modified FOTO Model**

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

**Table 5: Modified FOTO Backward Stepwise Model: Model summary & coefficients**

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 R2 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 5). Assumptions of normality and homoscedasticity were met and non-collinearity was satisfied. Sensitivity analysis showed that all variables that were predictive in OLS models remained significantly predictive in a mixed RE model. This model power remained with EQ5D5L used as an alternative PROM (adjusted R2 0.438) (see **Supplement 1**).

**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** alongside the standardised coefficients for the final backward stepwise model (modified NPROMs model version e) with all redundant/non-significant variables removed.

**Table 6: Modified NPROMs Backwards Stepwise Model: Model summary & coefficients**

Variables of IMD, sex, pain site and living alone were removed from the model due to not meeting statistical parameters (removed if p>0.1). ANOVA with these variables removed (model version e in **Table 6**) showed a large statistically significant F ratio (104.516, 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 R2 was 0.410 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. Sensitivity analysis showed that all variables that were predictive in OLS models remained significantly predictive in a mixed RE model. This model power remained with EQ5D5L used as an alternative PROM (adjusted R2 0.419) (see **Supplement 1**).

**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 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 7**).

Assumptions of normality and homoscedasticity were met and non-collinearity was satisfied.

Results show that adding in the variables with ‘moderate evidence’ did not statistically improve the model and therefore model version b using the variables with strong evidence to support their inclusion is the preferred model. This model explained 41% of the variation in 6-month MSK-HQ outcome and was statistically more predictive than model version a.

**Table 7: Keele Model 1 Summary: Evidence informed model**

**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).

Assumptions of normality and homoscedasticity were met and non-collinearity was satisfied. The model summary is shown in **Table 8**.

**Table 8: Keele Model 2: Statistically informed model (forward stepwise model demonstrating additional model strength as variables added)**

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 8**). 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 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 2 models tested the model based on variables within the FOTO case-mix methodology was slightly more predictive (R2 0.44) compared to the modified NPROMs model (R2 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 R2 0.37 (Deutscher et al, 2018) and NPROMs R2 0.23-0.30 NPROMs (DoH, 2012)). The predictive ability also remained high when an alternative PROM (EQ5D5L) was used (R2 0.44, R2 0.42 respectively). Both models removed variables of sex and pain site. Medication at baseline could not be added to the FOTO model in this study due to not being available at this time within the 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.

**Objective 2 Summary**

Keele Model 1 was developed using all variables with strong supporting evidence (model version b in **Table 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.

**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 in the FOTO model (Deutscher et al, 2018) 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, duration of symptoms, 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/body part, 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.

**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/test 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 R2 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.

**Conclusion**

In this paper four case-mix adjustment models are 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 setting on a large scale to further analyse the performance of this case-mix adjustment model, assessing its ability to benchmark performance including identification of variation in patient outcomes and care, and identification of optimal MSK care pathways. 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.

**References**

Baker, P. N., Deehan, D. J., Lees, D., Jameson, S., Avery, P. J., Gregg, P. J., & 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), 1058-1066.

Bryman, A., & Cramer, D. (2002). *Quantitative data analysis with SPSS release 10 for Windows: A guide for social scientists*. Routledge.

Burgess, R., Bishop, A., Lewis, M., & 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), 137-146.

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

Coles, (2010). PROMs risk adjustment methodology guide for general surgery and orthopaedic procedures. Northgate. <https://www.england.nhs.uk/statistics/wpcontent/uploads/sites/2/2013/07/proms-ris-adj-meth-sur-orth.pdf>

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

Deutscher, D., Werneke, M. W., Hayes, D., Mioduski, J. E., Cook, K. F., Fritz, J. M., ... & 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), 637-648.

EuroQol Research Foundation. EQ-5D-5L User Guide, 2019. Available from: https://euroqol.org/publications/user-guides.

Focus on Therapeutic Outcomes (FOTO) Website (2021). <https://www.fotoinc.com/>

Hart DL, Connolly JB. (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](http://www.cms.hhs.gov/TherapyServices/downloads/P4PFinalReport06-%2001-06.pdf).

Hill, J. C., Kang, S., Benedetto, E., Myers, H., Blackburn, S., Smith, S., ... & Price, A. (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).

Hill, J., Garvin, S., Chen, Y., Cooper, V., Wathall, S., Bartlam, B., ... & Foster, N. E. (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), 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., ... & Normand, S. L. T. (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), 456-462.

Lang, T. A., & 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), 5-9.

Lutz, A. D., Brooks, J. M., Chapman, C. G., Shanley, E., Stout, C. E., & 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), 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](https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/patient-reported-outcome-measures-proms/proms-clinical-case-study-data-informs-clinical-practice)

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>

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

Nuttall, D., Parkin, D., & 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), 41-54.

Price, A. J., Ogollah, R., Kang, S., Hay, E., Barker, K. L., Benedetto, E., ... & Hill, J. C. (2019). Determining responsiveness and meaningful changes for the Musculoskeletal Health Questionnaire (MSK-HQ) for use across musculoskeletal care pathways. *BMJ open*, *9*(10), e025357.

[Prince, (2012](https://www.sciencedirect.com/science/article/pii/S0165178119317834#bbib0037)). 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.Scott, I. C., McCray, G., Lancaster, G., Foster, N. E., & 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., & Cooray, T. M. J. A. (2019). Diagnosing multicollinearity of logistic regression model. *Asian Journal of Probability and Statistics*, 1-9.

Siemens, D. R., & Booth, C. M. (2017). Benchmarking our urological care: It’s just the beginning. *Canadian Urological Association Journal*, *11*(8), 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.