Received: 25 January 2023 | Accepted: 5 February 2023

DOI: 10.1002/msc.1744

WILEY

RESEARCH ARTICLE

Benchmarking quality of care using patient reported outcome measure data for patients presenting with musculoskeletal conditions in primary care GP practices

Roanna Burgess[1,2] | Martyn Lewis[1] | Jonathan C. Hill[1]

[1]Keele University, School of Medicine, Primary Care Centre Versus Arthritis, Staffordshire, UK

[2]Sandwell and West Birmingham Hospitals NHS Trust, Birmingham, UK

Correspondence
Roanna Burgess.
Email: r.m.burgess@keele.ac.uk

Abstract

Background: Patient Reported Outcome Measures (PROMs) evaluate health status from a patient perspective. They can be used to support care at a patient level but also collectively to review quality of care across care providers. Vast amounts of patients with musculoskeletal (MSK) conditions present to General Practice (GP) primary care practitioners each year. Variation in patient outcomes in this setting however has not been reported.

Objective: To identify variation in patient outcomes measured using the musculoskeletal health questionnaire (MSK-HQ) PROM for adults presenting to 20 GP practices in the UK with MSK conditions.

Methods: A secondary analysis of the STarT MSK cluster randomised controlled trial dataset. A standardised case-mix adjustment model, adjusting for condition complexity co-variates, was used to calculate predicted 6-month follow-up MSK-HQ scores, and used to compare adjusted and un-adjusted health gain (n = 868). Patient MSK-HQ change outcomes were aggregated to practice level and boxplots used to display outlier GP practices for un-adjusted and adjusted outcomes.

Results: Substantial variation in patient outcomes was seen across the 20 practices, even after case-mix adjustment, with mean change in MSK-HQ scores ranging from 6 to 12 points. Boxplots displaying un-adjusted outcomes showed one negative GP practice outlier and two positive outliers. However, the boxplots displaying case-mix adjusted outcomes showed no negative outliers, with two practices remaining as positive outliers, and one practice additionally becoming a positive outlier.

Conclusion: This study showed a two-fold GP practice variation in patient outcomes measured using the MSK-HQ PROM. To our knowledge it is the first study to demonstrate that (a) a standardised case-mix adjustment method can be used to fairly compare patient health outcome variation in GP care, and (b) that case-mix adjustment changes benchmarking findings with regards to provider performance and outlier identification. This has important implications for identifying best

practice exemplars and thereby helping to improve the quality of MSK primary care in the future.

KEYWORDS
musculoskeletal, outcome measures, primary health care

1 | INTRODUCTION

Patient Reported Outcome Measures (PROMs) consist of a series of questions that patients are asked to gauge views on their health, forming a self-assessment of a patient's health and health related quality of life (Devlin et al., 2010). They provide a standardised method for measuring patient's views (Ahmed, 2012) and therefore allow for collation and comparison. Clinicians use PROM data to guide clinical decision making but PROM data can also be used to evaluate comparative effectiveness when aggregated across patients (Van der Wees et al., 2014). Momentum among policy makers is growing for the routine and mandated collection and reporting of PROMs by clinical services. This standardised information can be used by commissioners and service leaders to aid decision making in relation to resource allocation and highlight best practice and variations in performance (Darzi, 2008). Until recently in the UK, there has been no universally agreed PROM for use in primary care or for use in musculoskeletal (MSK) practice. However, a recent consensus study (Burgess, Lewis, McRobert, & Hill, 2021) recommended the use of the musculoskeletal health questionnaire (MSK-HQ), pain intensity, and work absence and presenteeism for work age adults using the Work Productivity and Activity Index (Reilly et al., 1993) to evaluate MSK clinical outcomes in community and primary care settings, and these outcomes are now being routinely adopted within this setting.[1]

Comparing PROM scores for providers without taking into account patient complexity/provider case-mix can be misleading (Department of Health, 2012b) as the case-mix that one provider treats can be different to another. In order to make meaningful comparisons between providers a methodology is needed to make adjustment for these different provider profiles. These adjustments are normally based on statistical models which predict outcome, taking into account baseline clinical factors and characteristics which are beyond the providers' control (Department of Health, 2012b). This is commonly called 'case-mix' or 'risk' adjustment and allows for meaningful benchmarking of provider performance.

Benchmarking in clinical practice involves comparing and sharing best practice (Siemens et al., 2017). An example of successful MSK benchmarking within the UK, includes the National Health Service (NHS) England led National Patient Reported Outcomes (NPROMs) Programme for hip and knee arthroplasty (NHS Digital, 2022), with improvement methodology supported by the 'Getting it Right First Time' (GIRFT) Programme in Orthopaedics (GIRFT, 2020). The availability of data at a surgeon, unit, and trust level has allowed clinicians and managers to make informed decisions to improve the quality of patient care, including, improvement in the quality of implants used, reduction in low volume operating, and reduction in surgical site infections (GIRFT, 2020). So far these methods for identifying variation at unit and trust level have not been systematically applied outside of UK surgical settings.

Benchmarking can be based on either external criteria (normally based on research or judgement), or on an internal benchmark derived from the data itself, which is often the statistical mean of the data (over all records and providers of care) (Department of Health, 2012a). A threshold of three standard deviations from the mean is commonly used to identify values deemed 'out of control limits' (Department of Health, 2012a). This is equivalent to testing if a performance indicator is different from the target using a two-sided significance test at the significance level of 0.002 (control limit 99.8%) (Department of Health, 2012a). These limits are used in the UK National PROMs methodology to set 'alarms' for those providers falling outside of this limit (in the direction of worse health gain than the provider average) and to identify positive outliers (where care can be analysed to inform quality improvement initiatives based on best care). A threshold of two standard deviations from the mean, setting control limits at 95% (and significance level at 0.05) is used to identify 'alerts' (Department of Health, 2012a). Boxplots can also be used to identify and display outliers. Boxplots display low potential outliers (alert) when a score is more than 1.5 times the inter quartile range (IQR) but at most 3 times IQR below quartile 1 (25^{th} percentile), and high potential outliers when the score is more than 1.5 times IQR but at most 3 times IQR above quartile 3 (75^{th} percentile). Alarms (extreme outliers) are then scores which fall even further beyond these extremes (Ruben Geert van den Berg, 2022).

A previous study (Burgess, Lewis, & Hill, 2021) by the research team has shown that a modified US Focus on Therapeutic Outcomes (FOTO) case-mix adjustment model (Deutscher et al., 2018) is highly predictive of outcome in this STarT MSK Trial dataset ($R^2 = 0.44$). This was slightly more predictive than a model using the NPROMs case-mix adjustment model for MSK-HQ outcome data ($R^2 = 0.41$). For this study, we aimed to explore differences between un-adjusted and adjusted MSK-HQ health gain data using the modified FOTO case-mix adjustment model (adjusting for baseline clinical complexity) (Deutscher et al., 2018) and utilising the established NPROMs methodology for applying the case-mix model to a primary care MSK dataset (Department of Health 2012a, 2012b).

[1] Keele's National MSK Audit - Keele University

1.1 | Research question

How much variation in performance measured using the MSK-HQ PROM (health gain over 6 months) can be seen between General Practice (GP) practices who participated in the STarT MSK Trial, before and after adjustment for case-mix?

2 | METHOD

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 MSK conditions), cluster randomised controlled trial (RCT) in 2019/2020 (ISRCTN15366334 (Hill et al., 2020)). The trial did not show any significant differences in the patient outcomes of interest between the intervention and control arms and therefore the sample was treated as one cohort for the purposes of this analysis. A standardised set of metrics were collected for included patients (Hill et al., 2020), these included patient characteristics/demographics, baseline clinical factors, PROMs, and employment factors. The 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 statistical package Statistical Package for Social Sciences (SPSS) was used for all analyses.

The case-mix adjustment model adjusted data for baseline clinical complexity using pre-defined patient factors that are outside of provider/GP Practice control. The modified FOTO case-mix adjustment model as described by Burgess, Lewis, and Hill (2021) and originally developed by the FOTO team (Deutscher et al., 2018) was used to case-mix adjust patient outcomes (MSK-HQ) and includes the core variables of; baseline MSK-HQ score, previous pain episodes, comorbidities, current paid employment, previous surgery, physical activities, and duration of symptoms.

The case-mix adjustment analysis included data from the STarT MSK trial dataset for GP practices that had $n \geq 20$ patients with complete data (20 practices from original 24 practices had $n \geq 20$ with complete data). To ensure analyses were robust GP practices with less than 20 patients with complete data were excluded from the analysis (4 GP practices). The case-mix adjustment model was then applied to the dataset ($n = 868$), to calculate predicted 6-month follow up MSK-HQ scores for each patient.

Adjustment steps followed those described by NHS Digital's National PROMs method (see below) (NHS Digital, 2017). The case-mix adjustment process has three key stages: estimation of the impact of case-mix variables, generation of patient-level predicted scores, aggregation to organisation level and case-mix adjustment (NHS England, 2013).

1. Estimation of impact of case-mix variables (see Burgess, Lewis, & Hill, 2021, for full details of case-mix adjustment model development)
2. Generation of patient level predicted scores

The modified FOTO case-mix model previously described by Burgess, Lewis, and Hill (2021) was applied to the dataset using an ordinary least squares regression model to generate a predicted 6-month MSK-HQ follow up score for each patient.

3. Aggregation to organisation level and case-mix adjustment

The final step of 'aggregation' derives a final adjusted 6-month MSK-HQ follow up score for each organisation/practice to allow for fair comparison. This calculates the average difference between the expected values and the actual values at record level and adds this to the dataset average 6-month MSK-HQ follow up score (internal benchmark) (NHS Digital, 2017) (see formula A below based on reported formula by NHS Digital (2017) and adapted for this study).

(A) Adjusted MSK-HQ 6-month follow up score (Practice X) = Average MSK-HQ 6-month follow up score (STarT MSK trial dataset (38.61)) + Average (Actual MSK-HQ 6-month follow up score (Practice X)—Expected MSK-HQ 6-month follow up score (Practice X))

The adjusted health gain (the difference between baseline and 6-month MSK-HQ scores) was calculated as the difference between the adjusted MSK-HQ 6-month follow up score and the STarT MSK trial dataset average MSK-HQ baseline score (30.02) (see formula B below based on reported formula by NHS Digital (2017) and adapted for this study).

(B) Adjusted MSK-HQ Health Gain (Practice X) = Adjusted MSK-HQ 6-month follow up score (Practice X)—Average MSK-HQ baseline score (STarT MSK trial dataset (30.02))

If, on average, patients within a given GP practice scored themselves higher than their expected scores, that practice's adjusted score rose above the GP cohort average (for the 20 practices) and conversely, if the average score was lower than the average expected score, the practice's average fell below the GP cohort average (NHS Digital, 2017).

Following the above methodology:

1. Predicted scores were calculated at patient level within the dataset (calculated **predicted MSK-HQ 6-month follow up score** for each patient).
2. The difference between the actual score and the predicted score was then calculated at patient level.
3. Scores were then aggregated at Practice level to include **mean difference** in MSK-HQ scores at practice level between actual and predicted scores.

4. **Adjusted mean MSK-HQ 6-month follow up scores** were then calculated using formula A above, which involved taking the **average MSK-HQ 6-month follow up score** for the whole dataset (**38.61**) and adding it to the **mean difference** between actual and predicted scores at practice level to create an **adjusted mean MSK-HQ 6-month follow up score** per practice.
5. **Adjusted MSK-HQ health gain** was then calculated using formula B above, which involved taking the **adjusted mean MSK-HQ 6-month follow up score** for each practice and taking away the **mean baseline MSK-HQ score** for the dataset (**30.02**) to create the **adjusted MSK-HQ health gain** per practice.
6. Boxplots were then constructed for un-adjusted and adjusted MSK-HQ health gain across the 20 practices to display possible outliers and extreme values (these were used in place of funnel plots due to the small size of the dataset/number of practices (n = 20) allowing this to provide an easy to interpret visual of results including outlier practices).

3 | RESULTS

Table 1 displays descriptive statistics for the STarT MSK Trial full dataset and includes demographics, and all of the clinical severity factors included within the case-mix model. It also shows descriptive statistics for complete cases for included practices (with $\boldsymbol{n \geq 20}$ patients).

Table 2 displays the aggregated mean results at practice level. This table shows the practice code (1–20 GP practices), the **actual mean MSK-HQ 6-month follow up score** for each practice, the **predicted mean MSK-HQ 6-month follow up score** (predicted using the modified FOTO case-mix adjustment model using multiple linear regression), the **mean difference** between predicted and actual MSK-HQ 6-month follow up scores, the **adjusted mean MSK-HQ 6-month follow up scores** following formula A, and the **adjusted MSK-HQ health gain** at 6 months following formula B.

Table 3 shows descriptive statistics for un-adjusted and case-mix adjusted MSK-HQ health gain from baseline to 6-month follow up.

3.1 | Un-adjusted musculoskeletal health questionnaire health gain

Figure 1a and Figure 1b below show un-adjusted results for MSK-HQ health gain at 6 months (change in MSK-HQ from baseline to 6-month follow up) across the 20 practices (frequency gives number of practices).

It can be seen from the boxplot that there is one negative outlier identified from the 20 practices using un-adjusted data (Practice 14) and two positive outliers (Practice 1 and 7). None of these are at the 'extreme' level.

3.2 | Mean adjusted health gain for the musculoskeletal health questionnaire across the 20 GP practices

Table 3 shows descriptive statistics for adjusted MSK-HQ health gain from baseline to 6-month follow up.

Figures 2a and 2b below show case-mix adjusted results for MSK-HQ health gain at 6-month (change in MSK-HQ from baseline to 6-month follow up) across the 20 practices (frequency gives number of practices).

The boxplot in Figure 2b shows positive outliers for the cohort of 20 practices but no practices being a negative outlier when health gain is adjusted for case-mix and no 'extreme' outliers. Practices 1, 7, and additionally 12 are positive outliers when health gain is adjusted for case-mix.

When this is compared to un-adjusted MSK-HQ health gain scores there is agreement that Practices 1 and 7 are positive outliers. Practice 14 however is no longer a negative outlier, and practice 12 additionally becomes a positive outlier. None of these outliers are classed as extreme.

4 | DISCUSSION

Results show that for the 20 GP Practices, mean un-adjusted MSK-HQ scores at 6-month follow up showed substantial variation with mean scores ranging from 32.16 to 44.24 (a mean difference of 12.08 points). In comparison, the case-mix adjusted scores showed less variation with mean scores ranging from 35.86 to 41.78 (a mean difference of 5.92 points) showing that case-mix adjustment is meaningful and necessary. The mean un-adjusted health gain for patients within the 20 practices ranged from 2.08 to 13.93 (mean = 8.63, SD 2.57, mean difference 11.85), and when health gains were adjusted mean health gains ranged from 5.84 to 11.76 (mean = 8.72, SD 1.73, mean difference 5.92) (see Table 3). The results demonstrated significant variation in MSK-HQ patient outcome across different GP practices even after adjusting for baseline clinical complexity, with the best performing practices achieving over double the amount of MSK-HQ change on average per patient than the worst performing practice (and to a level equivalent to the size of the 6-point minimal clinically important difference (MCID) in the MSK-HQ (Price et al., 2019)). The majority of practices achieved the MCID of 6 (for average MSK-HQ health gain) both before and after case-mix adjustment. Two practices (Practices 14 and 20) did not meet this before adjustment for case-mix (10% of practices), reducing to only one practice (Practice 14) after case-mix adjustment (5% of practices).

There were three potential outliers in mean case-mix adjusted health gain across the 20 practices. Our study identified outliers using boxplots (Figures 1b and 2b), which helped to visualise outlier practices and showed how the results changed when case-mix adjustment was applied. After adjustment Practice 14 was no

TABLE 1 Descriptive statistics.

Descriptive statistics	Mean/frequency (%)	SD	*N*	% Complete data	Mean/frequency (%) for complete cases where practices had $n \geq 20$ patients	SD	*N*
N patients			1211	1211			868
N practices			24	24			20
MSK-HQ score baseline (mean)	29.17	10.162	1208	99.75	30.02	10.10	868
MSK-HQ score 6m Follow up (mean)	38.17	11.68	972	80.26	38.61	11.56	868
Age (mean)	60.03	15.28	1211	100	62.18	13.66	868
Sex (female) (*f* (%))	714 (59)		1211	100	501 (57.7)		868
Pain site (*f* (%))			1211	100			868
Knee	379 (31.3)				281 (32.4)		
Neck	130 (10.7)				93 (10.7)		
Back	457 (37.7)				317 (36.5)		
Shoulder	130 (10.7)				97 (11.2)		
Multisite	116 (9.6)				80 (9.2)		
Duration (without pain) (*f* (%))			1202	99.26			868
<3m	306 (25.5)				229 (26.4)		
3–6m	207 (17.2)				143 (16.5)		
7–12m	151 (12.6)				107 (12.3)		
1–2years	147 (12.2)				112 (12.9)		
3–5years	162 (13.5)				107 (12.3)		
6–10years	83 (6.9)				66 (7.6)		
Over 10years	146 (12.1)				104 (12.0)		
Previous pain episodes (*f* (%))			1205	99.50			868
0	260 (21.6)				195 (22.5)		
1	144 (12.0)				104 (12.0)		
2–3	231 (19.2)				162 (18.7)		
4–9	185 (15.4)				123 (14.2)		
10+	385 (32)				284 (32.7)		
Comorbidity count (*f* (%))			1211	100			868
0	386 (31.9)				257 (29.6)		
1	439 (36.3)				332 (38.2)		
2	250 (20.6)				180 (20.7)		
3 or more	136 (11.2)				99 (11.4)		
Physical activity (days per week) (*f* (%))			1208	99.75			868
0	299 (24.8)				204 (23.5)		
1	150 (12.4)				106 (12.2)		
2	177 (14.7)				126 (14.5)		
3	167 (13.8)				121 (13.9)		
4	105 (8.7)				76 (8.8)		

(Continues)

TABLE 1 (Continued)

Descriptive statistics	Mean/frequency (%)	SD	N	% Complete data	Mean/frequency (%) for complete cases where practices had $n \geq 20$ patients	SD	N
5	134 (11.1)				100 (11.5)		
6	42 (3.5)				31 (3.6)		
7	134 (11.1)				104 (12.0)		
Previous surgery (*f* (%))			1168	96.45			868
Yes	149 (12.2)				106 (12.2)		
No	1019 (87.8)				762 (87.8)		
Paid employment (*f* (%))			1165	96.20			868
Yes	561 (48.2)				392 (44.5)		
No	604 (51.8)				509 (56.5)		

TABLE 2 Aggregated mean musculoskeletal health questionnaire (MSK-HQ) 6-month follow up scores and mean MSK-HQ health gain for 20 GP practices before and after case-mix adjustment.

Practice	Actual MSK-HQ 6-month follow up score	Actual MSK-HQ health gain	Predicted MSK-HQ 6-month follow up score	Difference between actual and predicted MSK-HQ score	Adjusted MSK-HQ 6-month follow up score	Adjusted MSK-HQ health gain
1	40.52	13.93	37.32	3.17	41.78	11.76
2	39.78	8.04	39.47	0.31	38.92	8.9
3	37.09	9.04	37.26	−0.17	38.44	8.42
4	40.74	8.35	40.31	0.34	38.95	8.93
5	38.31	7.39	38.35	−0.04	38.57	8.55
6	37.58	8.88	38.15	−0.57	38.04	8.02
7	44.04	12.94	40.95	3.09	41.7	11.68
8	34.51	8	35.86	−1.36	37.25	7.23
9	39.15	9.15	38.93	0.22	38.83	8.81
10	35.44	8.51	35.99	−0.55	38.06	8.04
11	39.26	6.84	40.23	−0.98	37.63	7.61
12	42.96	11.54	39.84	3.12	41.73	11.71
13	44.24	10.97	41.92	2.32	40.93	10.91
14	33.96	2.08	36.71	−2.75	35.86	5.84
15	39.16	6.38	41.25	−2.09	36.52	6.5
16	40.74	10.26	39.67	1.07	39.68	9.66
17	40.64	8.21	40.24	0.4	39.01	8.99
18	37.48	8.25	37.56	−0.08	38.53	8.51
19	38.6	8.12	39.05	−0.45	38.16	8.14
20	32.16	5.81	34.5	−2.34	36.27	6.25

Note: (A) Adjusted MSK-HQ 6-month follow up score (Practice X) = Average MSK-HQ 6-month follow up score (STarT MSK trial dataset (38.61)) + Average (Actual MSK-HQ 6-month follow up score (Practice X)−Expected MSK-HQ 6-month follow up score (Practice X)).(B) Adjusted MSK-HQ Health Gain (Practice X) = Adjusted MSK-HQ 6-month follow up score (Practice X)−Average MSK-HQ baseline score (STarT MSK trial dataset (30.02)).

TABLE 3 Descriptive statistics showing un-adjusted and case-mix adjusted musculoskeletal health questionnaire (MSK-HQ) health gain at 6-month.

	Un-adjusted MSK-HQ health gain	Adjusted MSK-HQ health gain
N	20	20
Mean (95% CI)	8.63 (7.43–9.84)	8.72 (7.91–9.53)
Median	8.30	8.53
Standard deviation	2.57	1.73
Minimum	2.08	5.84
Maximum	13.93	11.76
Range	11.84	5.93
IQR	2.44	1.78

Abbreviations: CI, confidence interval; SD, standard deviation; IQR, inter quartile range.

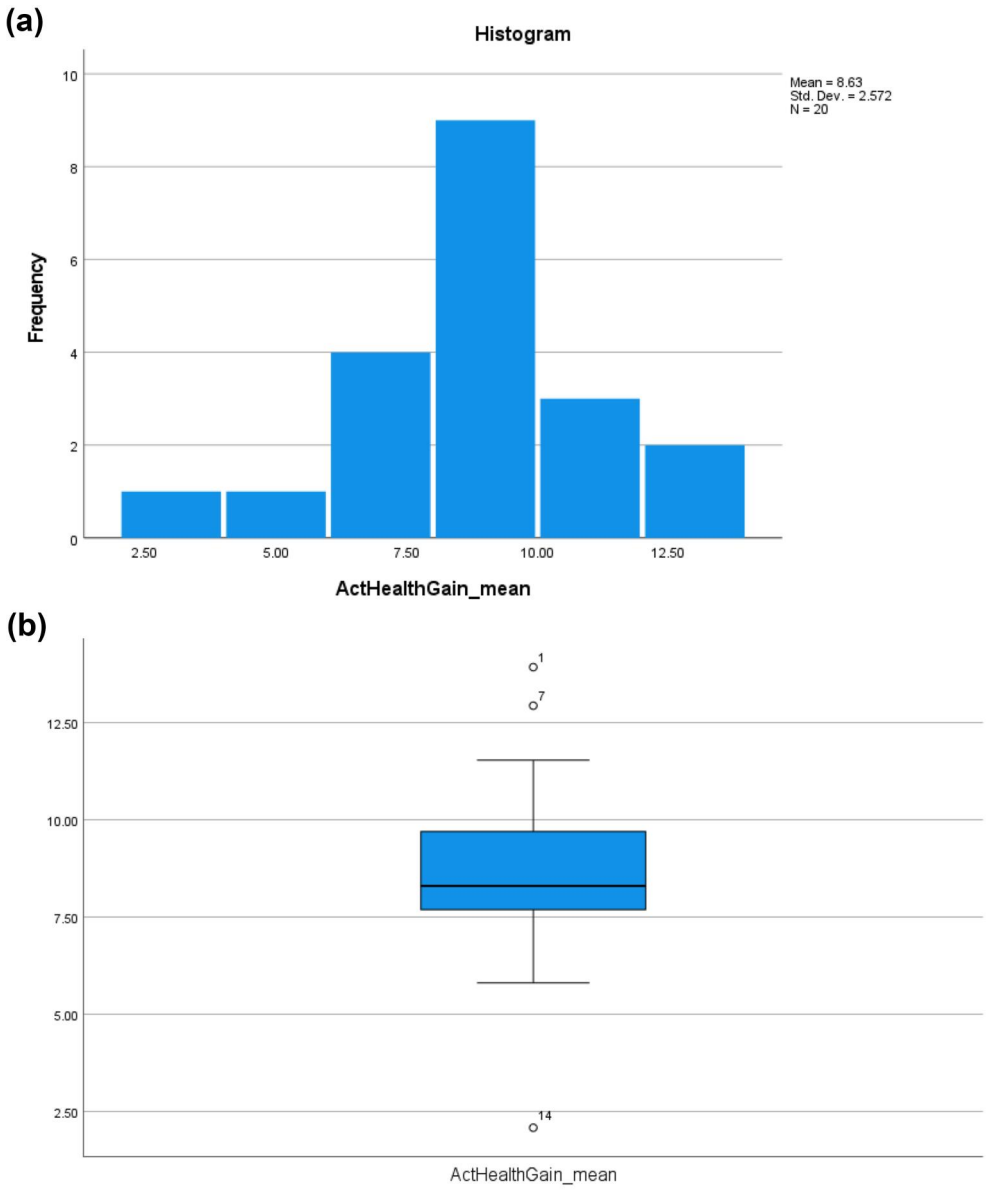

FIGURE 1 (a) Actual (un-adjusted) musculoskeletal health questionnaire (MSK-HQ) mean health gain for all 20 Practices. (b) Boxplot showing actual MSK-HQ mean health gain (un-adjusted) for 20 practices.

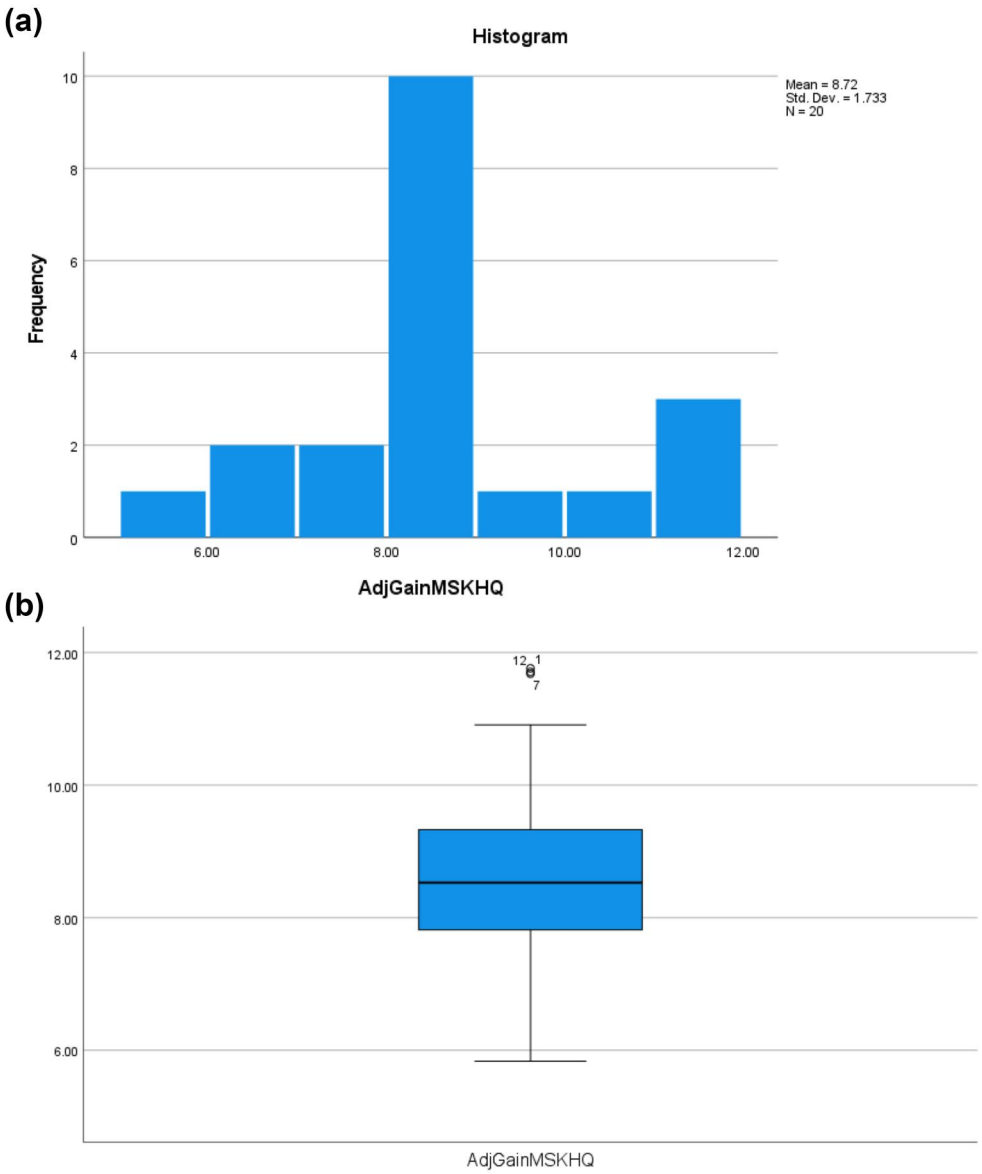

FIGURE 2 (a) Histogram showing adjusted musculoskeletal health questionnaire (MSK-HQ) mean health gain for all 20 Practices. (b) Boxplot showing adjusted MSK-HQ mean health gain for all 20 practices.

longer identified as a negative outlier, suggesting that the observed lower health gain in this GP practice was explained by baseline patient factors outside of Provider/Practice control (including for example, patients with: lower baseline MSK-HQ scores (indicating higher symptom severity), previous episodes of MSK pain, higher levels of comorbidities, no current employment, previous surgical history for their MSK condition, lower physical activity levels, and longer duration of symptoms). There were two positive outliers identified without adjustment (Practice 1 and Practice 7), which increased to three practices following case-mix adjustment (addition of Practice 12), as Practice 12 was only identified as a positive outlier after case-mix adjustment was applied. The provider differences explaining these three positive outliers in MSK outcomes are more likely to be due to factors that are within Provider/Practice control such as quality of care and warrant further investigation. For example, service organisational factors such as the size of practice, workforce configuration, access to physiotherapy or secondary care, or the quality of GP clinical decision making (e.g., referral rates, prescribing, investigations) could be viewed alongside benchmarked PROM data to understand what is different about the outlier practices that could be potentially driving better outcomes for patients. The research team plan to evaluate this further as part of the MIDAS study (led by Professor George Peat[2]), where PROM and patient reported experience measure data will be case-mix adjusted and benchmarked at aggregate level, and viewed alongside public health, organisational, and GP decision making metrics to gain a better understanding of variation across practices and what drives these

[2] MIDAS - Keele University

significant differences and inequities in patient outcomes and experiences.

4.1 | Limitations

Whilst this analysis provides an exemplar of how case-mix adjustment can be applied in a primary care setting for MSK patients to identify variation in outcome, it should be noted that the sample size was small ($n = 868$ with complete data) when compared to national datasets ($n = 12{,}084–15{,}718$ for individual orthopaedic models in the NPROMs developmental paper (Coles, 2010)) and therefore the results need to be viewed cautiously in this context. Within this analysis practices needed to have at least 20 complete patient records to be included, this is less than the 30 minimum number used for the NPROMs methodology (DoH, 2012b) however it was a pragmatic choice for the purposes of demonstration within this paper. Funnel plots were not used to identify outliers due to this smaller sample size. National Patient Reported Outcomes publications state that for provider organisations with volumes of less than 150, the use of symmetrical control limits may result in over-identification of negative outliers (DoH, 2012b). Further research is therefore needed with larger observational cohorts to confidently apply the NPROMs adjustment methods to identify outliers/variation in MSK outcomes across primary care. The case-mix model would also benefit from being externally validated in additional larger primary care observational cohorts to ensure all important confounders are included.

5 | CONCLUSION

This analysis demonstrated a two-fold variation in MSK patient outcomes across different GP practices, even when the results were adjusted for clinical case-mix/complexity. The study provides an example of how case-mix adjustment methodology can be applied in a primary care dataset to derive adjusted MSK-HQ follow-up scores and adjusted health gain at a provider (practice) level to identify performance outliers, using similar methods to the UK National PROMs methodology (NHS Digital, 2017). This data revealed clear positive outliers that could be further investigated as best practice exemplars, and after adjustment identified no practices performing significantly below the average and outside of control limits across the cohort. Next steps will include identifying the provider factors that were associated with positive outliers that seem to be driving above average performance. Learning realised from these types of analyses could be used to derive learning on best practice, to drive quality improvement initiatives in MSK patient outcomes, to help improve care within lower performing practices, as well as to help generate novel research questions.

AUTHOR CONTRIBUTION

All authors were involved in the writing and review of this paper.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest for this paper.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created in this study.

ETHICS STATEMENT

NA - This was a secondary analysis of an existing dataset. Data was collected within the STarT MSK (Subgrouping for Targeted Treatment in MSK conditions), cluster RCT in 2019/2020 (ISRCTN 15366334 (Hill et al., 2020)).

TWITTER

Roanna Burgess @BurgessRoanna

REFERENCES

Ahmed, S., Berzon, R. A., Revicki, D. A., Lenderking, W. R., Moinpour, C. M., Basch, E., Reeve, B. B., & Wu, A. W. (2012). The use of patient-reported outcomes (PRO) within comparative effectiveness research: Implications for clinical practice and health care policy. *Medical Care*, *50*(12), 1060–1070. https://doi.org/10.1097/mlr.0b013e318268aaff

Burgess, R., Lewis, M., & Hill, J. C. (2021). Musculoskeletal case-mix adjustment in a UK primary/community care cohort: Testing musculoskeletal models to make recommendations in this setting. *Musculoskeletal Science and Practice*, *56*, 102455. https://doi.org/10.1016/j.msksp.2021.102455

Burgess, R., Lewis, M., McRobert, C., & Hill, J. C. (2021). Developing a core outcome set for community and primary care musculoskeletal services: A consensus approach. *Musculoskeletal Science and Practice*, *55*, 102415. https://doi.org/10.1016/j.msksp.2021.102415

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

Darzi, A. (2008). High quality care for all: NHS next stage review final report. Cm 7432. London: The stationery office. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/228836/7432.pdf

Deutscher, D., Werneke, M. W., Hayes, D., Mioduski, J. E., Cook, K. F., Fritz, J. M., Woodhouse, L. J., & 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. https://doi.org/10.2519/jospt.2018.7981

Devlin, N., & Appleby, J. (2010). *Getting the most out of PROMs: Putting health outcomes at the heart of NHS decision making*. The King's Fund. https://www.kingsfund.org.uk/sites/default/files/Getting-the-most-out-of-PROMs-Nancy-DevlinJohn-Appleby-Kings-Fund-March-2010.pdf

DoH. (2012a). Patient reported outcome measures (PROMs) in England A methodology for identifying potential outliers. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/216250/dh_133579.pdf

DoH. (2012b). Patient reported outcome measures (PROMs) in England: The case-mix adjustment methodology. Patient Reported Outcome Measures (PROMs) in England: The case-mix adjustment methodology. Retrieved from publishing.service.gov.uk

Geert van den Berg, R. (2022). SPSS tutorials website. Available at. Boxplots - Beginners Tutorial with Examples spss-tutorials.com

GIRFT. (2020). Getting it Right in Orthopaedics: Reflecting on success and reinforcing improvement. A follow-up on the GIRFT national specialty report on orthopaedics. Available at: https://gettingitrightfirsttime.co.uk/wp-content/uploads/2020/02/GIRFT-orthopaedics-followup-report-February

Hill, J., Garvin, S., Chen, Y., Cooper, V., Wathall, S., Bartlam, B., Saunders, B., Lewis, M., Protheroe, J., Chudyk, A., Birkinshaw, H., Dunn, K. M., Jowett, S., Oppong, R., Hay, E., van der Windt, D., Mallen, C., & 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. https://doi.org/10.2196/17939

Hill, J. C., Kang, S., Benedetto, E., Myers, H., Blackburn, S., Smith, S., Dunn, K. M., Hay, E., Rees, J., Beard, D., Glyn-Jones, S., Barker, K., Ellis, B., Fitzpatrick, R., & 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), e012331. https://doi.org/10.1136/bmjopen-2016-012331

NHS Digital. (2017). Patient reported outcome measures (PROMs) in England A guide to PROMs methodology. Retrieved from https://webarchive.nationalarchives.gov.uk/ukgwa/20180307185917tf_/http://content.digital.nhs.uk/media/1537/A-Guide-to-PROMs-Methodology/pdf/PROMs_Guide_V12.pdf

NHS Digital. (2022). Finalised patient reported outcome measures (PROMs) in England, for hip and knee replacement procedures (April 2020 to March 2021). Available at: https://digital.nhs.uk/data-and-information/publications/statistical/patient-reported-outcome-measures-proms/finalised-hip-and-knee-replacement-procedures-april-2020-to-march-2021#:~:text=95%25%20of%20hip%20replacement%20patients,87%25%20of%20knee%20replacement%20patients

NHS England. (2013). Patient reported outcome measures in England: Update to reporting and case-mix adjusting hip and knee procedure data. Available at: https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2013/10/proms-meth-prim-revis.pdf

Price, A. J., Ogollah, R., Kang, S., Hay, E., Barker, K. L., Benedetto, E., Smith, S., Smith, J., Galloway, J. B., Ellis, B., Rees, J., Glyn-Jones, S., Beard, D., Fitzpatrick, R., & 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. https://doi.org/10.1136/bmjopen-2018-025357

Reilly, M. C., Zbrozek, A. S., & Dukes, E. M. (1993). The validity and reproducibility of a work productivity and activity impairment instrument. *PharmacoEconomics*, *4*(5), 353–365. https://doi.org/10.2165/00019053-199304050-00006

Scott, I. C., McCray, G., Lancaster, G., Foster, N. E., & Hill, J. C. (2020). Validation of the Musculoskeletal Health Questionnaire (MSK-HQ) in primary care patients with musculoskeletal pain*Seminars in Arthritis and Rheumatism*. *50*(5), pp. 813–820. WB Saunders.

Siemens, D. R., & Booth, C. M. (2017). Benchmarking our urological care: It's just the beginning. *Canadian Urological Association Journal*, *11*(8), 223–224. https://doi.org/10.5489/cuaj.4803

Van der Wees, P. J., Nijhuis-van Der Sanden, M. W., Ayanian, J. Z., Black, N., Westert, G. P., & Schneider, E. C. (2014). Integrating the use of patient-reported outcomes for both clinical practice and performance measurement: Views of experts from 3 countries. *The Milbank Quarterly*, *92*(4), 754–775. https://doi.org/10.1111/1468-0009.12091

How to cite this article: Burgess, R., Lewis, M., & Hill, J. C. (2023). Benchmarking quality of care using patient reported outcome measure data for patients presenting with musculoskeletal conditions in primary care GP practices. *Musculoskeletal Care*, 1–10. https://doi.org/10.1002/msc.1744