

ORIGINAL ARTICLE

Minimal important changes and differences were estimated for Oxford hip and knee scores following primary and revision arthroplasty

Shiraz A. Sabah^{a,*}, Abtin Alvand^{a,b}, David J. Beard^a, Andrew J. Price^{a,b}^aNuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford^bNuffield Orthopaedic Centre, Oxford

Received in revised form 6 December 2021; Accepted 9 December 2021; Available online 15 December 2021

Abstract

Objectives: To create estimates for clinically meaningful changes and differences in pain and joint function for the Oxford Hip and Knee Scores (OHS/OKS) in primary and revision joint replacement.

Study Design and Setting: 694,487 primary and revision joint replacement procedures were analyzed from the NHS PROMs dataset between 2012–2020. Minimal important changes (MIC) and differences (MID) were calculated using distribution and anchor-based methods (including receiver-operating characteristic (ROC) curve and predictive-modelling techniques).

Results: For comparison of two or more groups (such as in a clinic trial), MID estimates were ~5 points. For cohort studies investigating changes over time in a single group of patients, MIC_{group} estimates were 12.4 points (primary hip replacement), 8.6 points (revision hip replacement), 10.5 points (primary knee replacement) and 9.4 points (revision knee replacement). For studies investigating changes over time at the individual patient level, $MIC_{adjusted}$ estimates were ~8 points, ~6 points, ~7 points and ~6 points respectively.

Conclusion: This study has calculated contemporary estimates of clinically important changes and differences for the OHS/OKS for primary and revision hip and knee replacement. These estimates can be used to inform sample size calculations and to interpret changes in joint function over time and differences between groups. © 2021 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

Keywords: Total hip replacement; Total knee replacement; Patient-reported outcome measures; Revision; Re-operation

1. Introduction

Patient-reported outcome measures (PROMs) have established roles in the measurement of outcome in randomised controlled trials [2,3] and observational studies [4], as well as more contemporary roles comparing health-care delivery between providers [5] and supporting patient decision-making [6]. In the UK, the collection of PROMs before and after hip and knee replacement surgery has been

mandated since 2009 as part of a national audit [5]. This programme uses two well-established PROM instruments, the Oxford Hip Score (OHS) and the Oxford Knee Score (OKS) [7,8], to evaluate pain and joint function during activities of daily living [9,10]. Both instruments have been found to have good measurement properties and evidence to support their use in primary joint replacement [11]. The instruments are also valid for use in revision hip and knee joint replacement procedures where the goal of surgery is the alleviation of pain or improvement in joint function [12,13].

All PROMs must be evaluated to determine their measurement properties, such as their reliability and validity [14]. The interpretability of a PROM refers to the ability to assign qualitative meaning to the score itself, a change in the score over time or a difference in score between groups [14]. This concept is useful when judging the efficacy of joint replacement, where one needs to be able to decide whether changes or differences in a PROM score are clinically meaningful [15]. For cohort studies,

Conflict of interests: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding: Mr. Shiraz Sabah received funding from the Royal College of Surgeons 1-Year Fellowship, Rosetrees Trust and National Institute of Health Research (NIHR). None of the funding sources had any role in the design or conduct of the study.

Data statement: All data for this study is available within the public domain on the NHS Digital PROMs website and is regularly updated [1]. Statistical code is provided as a supplementary file.

* Corresponding author. Tel +44 1865 223421.

E-mail address: shiraz.sabah@ndorms.ox.ac.uk (S.A. Sabah).

What is new?

- We have provided new estimates for clinically meaningful changes and differences for the Oxford Hip and Knee Scores (OHS/OKS) in primary and revision total hip and knee replacement (THR/TKR).
- This study used a large patient sample from the NHS PROMs dataset and upgraded psychometric methodology to create precise estimates that were corrected for the high prevalence of improved patients.
- Future RCTs investigating pain and function after THR/TKR can use a minimal important difference (MID) of 5 points for sample size calculations.

the minimal important change (MIC) has been defined as the smallest change observed over a period of time in a single individual or group that can be considered to be clinically meaningful [15]. For comparative studies (such as randomized controlled trials), the concept of the minimal important difference (MID) is defined as the smallest difference in score between two or more groups at the time of outcome assessment that is considered to be clinically meaningful [15].

We have previously produced MIC and MID estimates for primary hip and knee replacement using early data from the NHS PROMs programme using established anchor-based methodology [15]. However, new methodology for calculating MICs has emerged that may offer more precise [16] and less biased [17] estimates using regression-based techniques. These developments impact upon the correct use of estimated MIC values and regular re-evaluation is required. In addition, MIC and MID estimates for revision joint replacement surgery have not been calculated.

As such, the aim of this study was to investigate and create estimates for meaningful changes and differences in joint function following primary and revision hip and knee joint replacement using the latest available national data and upgraded psychometric methodology.

2. Materials and methods

This study used publicly available, de-identified data from the NHS PROMs programme, which does not require ethical approval. The dataset included responses to PROM questionnaires for publicly-funded, elective primary and revision hip and knee replacement procedures performed in England from first April 2012 to 31 March 2020. Further information on this dataset is available online from NHS Digital [1]. Instructions and statistical code for downloading, cleaning and processing this data are provided as a supplementary file.

2.1. Outcome measures

The OHS and OKS measure patient-reported pain and function over a 4-week period using twelve Likert items [7,8,18]. Each item is rated from zero to four, to give a best possible total score of 48 points. For the national audit, patients complete a pre-operative questionnaire, followed by a second questionnaire administered at 6 months following surgery. The follow-up questionnaire includes an anchor question used to record perceived success following surgery: “Overall, how are the problems now in the hip/knee on which you had surgery, compared to before your operation?” Response: *Much better, A little better, About the same, A little worse, Much worse.*

2.2. Statistical analysis

2.2.1. Characteristics of the study population

Primary and revision hip and knee replacements were each analysed separately. Each study population was described using frequencies, means and standard deviations or medians and interquartile ranges as appropriate. We calculated meaningful changes and differences based on patients with complete data (which we defined as valid responses to each OHS or OKS item and the anchor question for perceived success).

2.2.2. Appropriateness of the anchor question to measure change in the PROM score

The appropriateness of the anchor question to measure change in the PROM score was evaluated using Spearman’s rank correlation coefficients. The strength of correlation between the transition item and the PROM change score was interpreted using the credibility instrument developed by Devji et al [19] as a reference. An absolute correlation coefficient of the anchor question to the change score of ≥ 0.5 was interpreted as *high credibility*. To assess for dependency of the anchor on the baseline or final status, the anchor question was further correlated to the pre-operative and post-operative PROM scores and an assessment of *high* or *low credibility* was made by following guides within the credibility instrument [19].

2.2.3. Estimates for minimal important change (MIC)

The minimal important change (MIC) can be understood as the smallest change in a PROM score over a period of follow-up that is interpreted by a single patient or group to be clinically meaningful. Two main approaches are used to estimate this quantity. *Distribution-based* methods report statistical properties of the sample, such as the spread of observed data (e.g. Standard Error of Measurement, SEM) or its ‘signal-to-noise ratio’ (e.g. Effect size, ES). Whilst these estimates do not necessarily indicate clinical importance, some have argued that simple rules can be generalised across these measures to allow this – for example, the observation that one half standard deviation of

difference is a useful threshold to discriminate many conditions [20]. *Anchor-based* methods reference changes or differences in the PROM score to an external criterion to which the patient has placed importance, such as a global transition item (which itself must have validity and be interpretable) [21] and so are preferred by the CONsensus-based Standards for selection of health status Measurement INstruments (COSMIN) group [14].

2.2.4. Distribution-based methods

Distribution-based methods were selected for reporting here if they produced estimates in the same units as the PROM (i.e. a change in points). Estimates without units (such as the standardized effect size or standardized response mean) were not calculated.

2.2.4.1. Half standard deviation. This is half the standard deviation of the PROM score at baseline, which has been shown to be a good approximation for many estimates of clinically important changes [20].

2.2.4.2. Standard error of measurement (SEM). This estimate incorporates the reliability of the measurement instrument and is calculated as: $sd_{baseline} \times \sqrt{(1 - reliability)}$. Cronbach alpha [22] was used to provide the reliability estimate, which aligns with the interpretation of Dawson et al [23] that the SEM represents “the error estimate for a single use of the questionnaire”.

2.2.4.3. Minimal detectable change (MDC_{90}). This represents the smallest change score that can be considered beyond the measurement error of the instrument. It is calculated by multiplying the SEM by $\sqrt{2}$ to account for the error introduced through measurement on two occasions and a z value representing the desired confidence level for the uncertainty of the observed score (with 1.65 representing a 90% confidence level) [15].

2.2.5. Anchor-based methods

2.2.5.1. MIC_{group} . This is the minimal important change for improvement (MIC) for a single *group* of patients over time. It is calculated from the mean change score in the group that responded ‘a little better’ to the anchor question. 95% confidence intervals were estimated using *Rmisc* package in R [24].

2.2.5.2. MIC_{ROC} . This is the MIC for a single *individual* over time. It is calculated using receiver operating characteristic (ROC) curves to discriminate between patients who responded to indicate the smallest, meaningful clinical improvement (‘a little better’) versus those who responded to indicate no change (‘about the same’). The best cut-point (Youden index) was defined as the maximum of sensitivity and specificity using the R package *cutpointr* [25]. 95% confidence intervals were estimated using bootstrap replications ($n = 1000$).

2.2.5.3. MIC_{pred} . The MIC_{pred} is a predictive modelling approach to estimation of a minimal important change. It has been proposed as a more precise method than the MIC_{ROC} and also allows for the incorporation of effect modifiers. An excellent description of the MIC_{pred} can be found in the article by Terluin et al [16]. To calculate the MIC_{pred} , we constructed a logistic regression model to predict group membership. The anchor question was used as the outcome variable. Those patients who responded ‘a little better’ were considered to be minimally importantly improved, whilst those patients who responded ‘about the same’ were considered to be unchanged from baseline. Patients with responses of ‘much better’, ‘a little worse’ or ‘much worse’ were excluded. The predictor variable was the change in OHS/OKS. This provided the logistic regression equation:

$$\ln(odds_{post}) = C + B_x * x$$

where C represented the intercept and B_x represented the regression coefficient of the change in Oxford Hip/Knee Score, x . The $odds_{post}$ provided the odds of belonging to the minimally importantly improved group for a given change in Oxford Hip/Knee Score.

The Likelihood ratio (LR) was used to represent the post-test odds of belonging to the improved group divided by the pre-test odds, i.e.:

$$LR = odds_{post} / odds_{pre}$$

A LR of 1 was used to represent the MIC_{pred} .

2.2.5.4. $MIC_{adjusted}$. A subsequent study by Terluin et al [17] recommended adjustment of the MIC_{pred} based on the proportion of improved patients. They suggested the following formula:

$$MIC_{adjusted} = MIC_{pred} - (0.090 + 0.103 * Cor) * sd_{change} * \log - odds(imp)$$

where Cor is the point biserial correlation between the change in OHS/OKS and the anchor, sd_{change} is the standard deviation of the change in OHS/OKS; and $\log - odds(imp)$ is the natural logarithm of (proportion improved / (1 - proportion improved)). The theoretical basis of this formula is not clear [17].

2.3. Estimates for minimal important difference (MID)

The minimal important difference (MID) was defined as the smallest difference in PROM scores between groups at the time of outcome assessment that was interpreted to be clinically meaningful. The MID was calculated as the difference in the mean change score for patients who responded ‘a little better’ compared to those who responded ‘about the same’ to the anchor question.

2.2.3. Software

Statistical analyses were performed using R version 4.1.2.

(a) Hips



(b) Knees

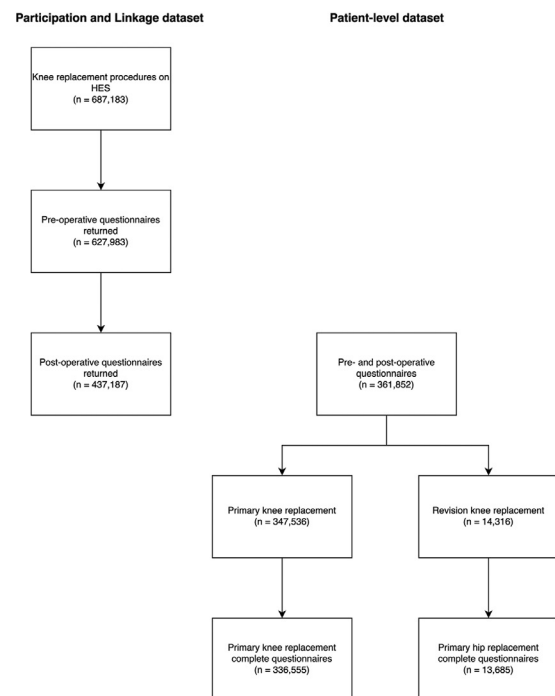


Fig. 1. Flowchart demonstrating patient participation in the NHS PROMs programme and attrition of records for (A) hip replacement and (B) knee replacement procedures.

3. Results

3.1. Characteristics of the study population

A study flowchart is provided in Fig. 1. The study population with complete data from which meaningful changes and differences were calculated represented ~50% of the total number of procedures performed (hip replacement $n = 321238/618358$ [52%]; knee replacement $n = 350240/687183$ [51%]). The baseline characteristics of patients undergoing primary and revision hip and knee replacement surgery are summarised in Table 1. For each of the procedures evaluated, there were more females than males and the most frequent age group was 70–79 years. Post-operative PROM scores, change scores and perceived success were higher following primary joint replacement compared to revision joint replacement. Each of the pre-operative scores was approximately normally distributed and the post-operative scores were negatively skewed (Fig. 2).

3.2. Appropriateness of the anchor question to measure change in the PROM score

The correlation coefficients for the anchor question with change, pre-operative and post-operative scores are shown in Table 2. The core criterion of whether the anchor question had good correlation with the PROM change score found ‘high credibility’ for all procedures, except for primary hip replacement. The anchor question correlated ‘to

a great extent’ with the PROM score at baseline for all procedures except for revision knee replacement (‘not so much’). For all procedures, the anchor question demonstrated greater correlation to the post-operative score than to the change score, which may indicate recall bias. The relationship between the PROM change score and each anchor question item is presented in Fig. 3.

3.3. Estimates for minimal important change (MIC)

MIC estimates calculated using distribution- and anchor-based methods are summarised in Table 3. We found that estimates for minimal important change on a group-level (MIC_{group}) had greater magnitude than estimates on an individual patient level. A single group of patients or an individual undergoing revision joint replacement required a smaller increase in their PROM score to recognise a meaningful improvement compared to their counterpart undergoing primary joint replacement.

3.3.1. Primary hip replacement

• Distribution-based estimates

The half standard deviation estimate was ± 4.1 points. The SEM estimate was ± 2.6 points. The MDC_{90} estimate was ± 6 points.

• Anchor-based estimates

The mean change in OHS for patients responding ‘a little better’ was 12.4 points (95% CI 12.3 – 12.5). The MIC_{ROC} estimate was 9 points (95% CI 9 – 10) with

Table 1. Patient characteristics.

Characteristic	Primary hip replacement N = 313,338	Revision hip replacement N = 19,297	Primary knee replacement N = 347,536	Revision knee replacement N = 14,316
Age				
<60 years	41,568 (13%)	2,199 (11%)	34,146 (9.8%)	1,657 (12%)
60 - 69 years	97,304 (31%)	5,117 (27%)	116,968 (34%)	4,696 (33%)
70 - 79 years	114,845 (37%)	7,555 (39%)	136,855 (39%)	5,336 (37%)
80 + years	40,296 (13%)	3,457 (18%)	40,870 (12%)	1,782 (12%)
Not specified	19,325 (6.2%)	969 (5.0%)	18,697 (5.4%)	845 (5.9%)
Gender				
Male	114,683 (37%)	7,909 (41%)	139,930 (40%)	6,374 (45%)
Female	179,256 (57%)	10,419 (54%)	188,802 (54%)	7,097 (50%)
Not specified	19,399 (6.2%)	969 (5.0%)	18,804 (5.4%)	845 (5.9%)
Oxford Score				
Pre-operation	18 (8)	21 (11)	19 (8)	16 (8)
Missing	3,445	300	4,012	234
Post-operation	40 (9)	34 (11)	36 (9)	29 (11)
Missing	3,740	355	5,794	315
Change in score	22 (10)	13 (12)	17 (10)	12 (11)
Missing	7,079	631	9,680	533
Perceived success				
Much better	270,697 (86%)	12,612 (65%)	256,150 (74%)	7,559 (53%)
A little better	25,326 (8.1%)	3,066 (16%)	53,555 (15%)	2,981 (21%)
About the same	6,604 (2.1%)	1,554 (8.1%)	15,072 (4.3%)	1,442 (10%)
A little worse	3,915 (1.2%)	902 (4.7%)	11,549 (3.3%)	1,089 (7.6%)
Much worse	2,850 (0.9%)	713 (3.7%)	7,462 (2.1%)	1,000 (7.0%)
Not specified	3,946 (1.3%)	450 (2.3%)	3,748 (1.1%)	245 (1.7%)

n(%); mean(sd)

Table 2. Correlation matrix for the success transition question and OHS/OKS.

Success transition versus ...	Primary hip replacement <i>rho</i> Credibility rating		Revision hip replacement <i>rho</i> Credibility rating		Primary knee replacement <i>rho</i> Credibility rating		Revision knee replacement <i>rho</i> Credibility rating	
Change score ^{*,†}	-0.42	Low	-0.54	High	-0.55	High	-0.68	High
Pre-operative score ^{*,†}	-0.03	High	-0.04	High	-0.09	High	-0.14	Low
Post-operative score ^{*,†,‡}	-0.46	Low	-0.57	Low	-0.60	Low	-0.72	Low

* all $P < 0.0001$

† Correlation coefficients interpreted using guides from the credibility rating instrument developed by Devji et al.

‡ Higher correlation of the transition item to the post-operative score compared to the change score was rated low credibility.

Table 3. Summary table of MIC estimates for oxford hip and knee scores.

Distribution-based estimates	Primary hip replacement	Revision hip replacement	Primary knee replacement	Revision knee replacement
Half standard deviation	4.1	5.5	3.9	4
Standard error of measurement	2.6	2.8	2.6	2.7
Minimal detectable change-90	6	6.6	6	6.3
Anchor-based estimates				
MIC-group	12.4 (12.3 - 12.5)	8.6 (8.3 - 8.9)	10.5 (10.5 - 10.6)	9.4 (9.1 - 9.7)
MIC-ROC	9 (9 - 10)	6 (5 - 10)	9 (9 - 9)	8 (6 - 8)
MIC-PRED	8.9 (8.3 - 9.5)	6.5 (5.3 - 7.8)	8.1 (7.6 - 8.5)	6.8 (5.6 - 8.2)
MIC-ADJ	7.6 (7 - 8.2)	5.8 (4.6 - 7.1)	6.9 (6.5 - 7.3)	6.2 (5 - 7.6)

MIC, minimal important change; group, mean change method; ROC, receiver operating characteristic curve method; PRED, predictive modelling method; ADJ, adjusted estimate following predictive modelling based on prevalence of improvers

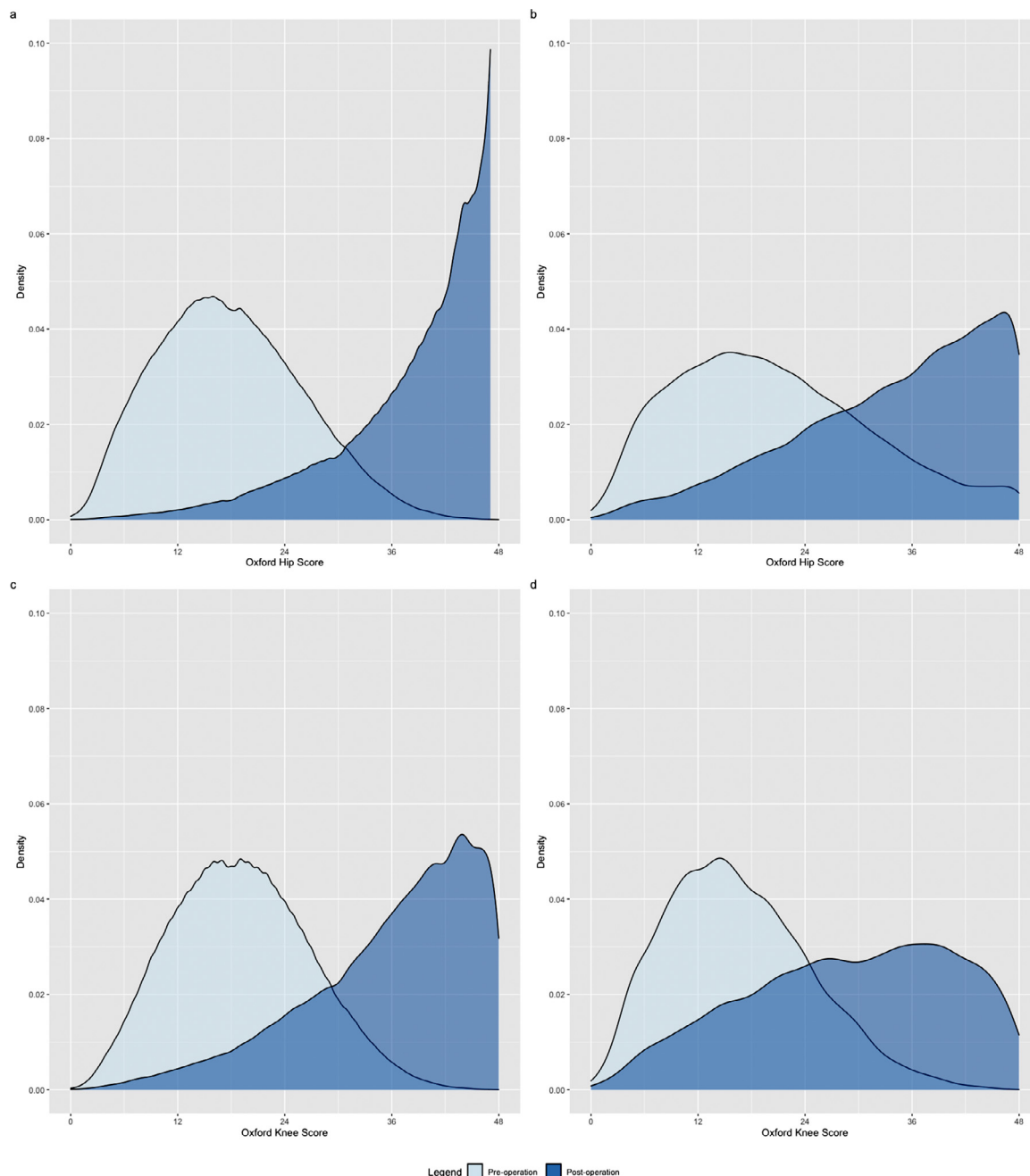


Fig. 2. Kernel Density plots of pre- versus post-operative OHS/OKS for (A) primary hip replacement, (B) revision hip replacement, (C) primary knee replacement and (D) revision knee replacement.

sensitivity 0.68, specificity 0.58 and area under the curve (AUC) 0.67. The proportion improved was 0.79. The MIC_{pred} estimate was 8.9 points (95% CI 8.3 - 9.5). The $MIC_{adjusted}$ estimate was 7.6 points (95% CI 7 - 8.2).

3.3.2. Revision hip replacement

• Distribution-based estimates

The half standard deviation estimate was ± 5.5 points. The SEM estimate was ± 2.8 points. The MDC_{90} estimate was ± 6.6 points.

• Anchor-based estimates

The mean change in OHS for patients responding ‘a little better’ was 8.6 points (95% CI 8.3 - 8.9). The MIC_{ROC} estimate was 6 points (95% CI 5 - 10) with sensitivity 0.65, specificity 0.61 and area under the curve (AUC) 0.67. The proportion improved was 0.66. The MIC_{pred} estimate was 6.5 points (95% CI 5.3 - 7.8). The $MIC_{adjusted}$ estimate was 5.8 points (95% CI 4.6 - 7.1).

3.3.3. Primary knee replacement

• Distribution-based estimates

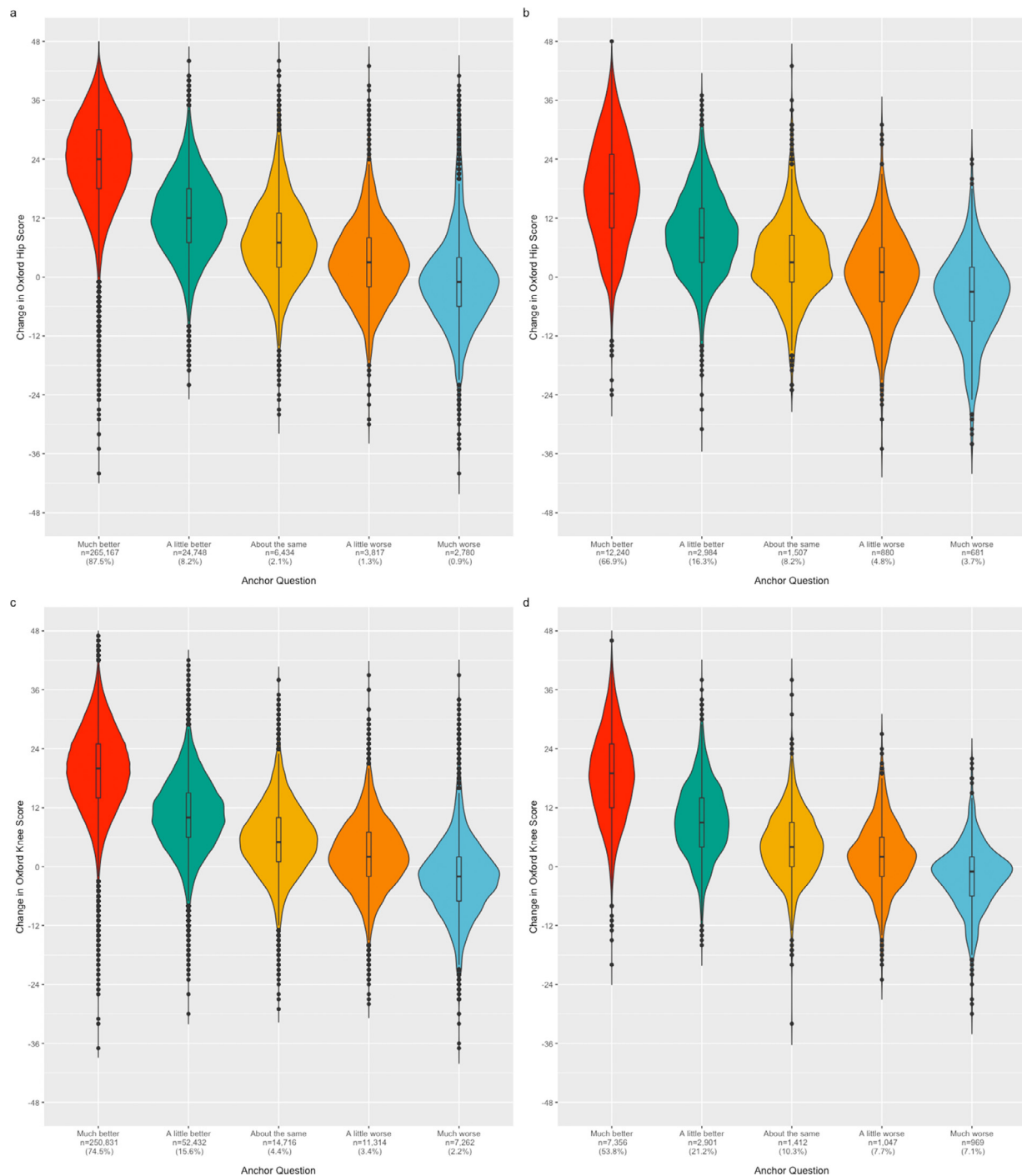


Fig. 3. Violin-plots demonstrating the relationship of the anchor question to the change in OHS/OKS for (A) primary hip replacement, (B) revision hip replacement, (C) primary knee replacement and (D) revision knee replacement.

The half standard deviation estimate was ± 3.9 points. The SEM estimate was ± 2.6 points. The MDC_{90} estimate was ± 6 points.

- Anchor-based estimates

The mean change in OKS for patients responding ‘a little better’ was 10.5 points (95% CI 10.5 - 10.6). The MIC_{ROC} estimate was 9 points (95% CI 9 - 9) with sensitivity 0.61, specificity 0.68 and area under the curve (AUC)

0.69. The proportion improved was 0.78. The MIC_{pred} estimate was 8.1 points (95% CI 7.6 - 8.5). The $MIC_{adjusted}$ estimate was 6.9 points (95% CI 6.5 - 7.3).

3.3.4. Revision knee replacement

- Distribution-based estimates

The half standard deviation estimate was ± 4 points. The SEM estimate was ± 2.7 points. The MDC_{90} estimate was ± 6.3 points.

- Anchor-based estimates

The mean change in OKS for patients responding ‘a little better’ was 9.4 points (95% CI 9.1 - 9.7). The MIC_{ROC} estimate was 8 points (95% CI 8 - 8) with sensitivity 0.6, specificity 0.71 and area under the curve (AUC) 0.7. The proportion improved was 0.67. The MIC_{pred} estimate was 6.8 points (95% CI 5.6 - 8.2). The $MIC_{adjusted}$ estimate was 6.2 points (95% CI 5 - 7.6).

3.4. Estimates for minimal important difference (MID)

The MID estimates (calculated as the difference in mean scores between patients who responded ‘a little better’ compared to those who responded ‘about the same’) were each ~5 points (4.9 points primary hip replacement, 5 points for revision hip replacement, 5 points for primary knee replacement, and 5.1 points for revision knee replacement).

4. Discussion

This study has investigated clinically important changes and differences in the Oxford Hip and Knee Scores. These are updated estimates for primary joint replacement surgery and new estimates for revision surgery. For each of these procedures, the MID estimate was ~5 points. This estimate can be used to inform sample size calculations for clinical trials using these instruments in the future. For cohort studies investigating a single group of patients over time, we found that an increase in OHS of 12.4 points was needed on average to interpret improvement after primary hip replacement and 8.6 points for revision hip replacement. The corresponding MIC_{group} estimates for the OKS after primary and revision knee replacement were 10.5 points and 9.4 points, respectively. The estimates for meaningful change were smaller when interpreted on an individual patient level. The $MIC_{adjusted}$ method produced estimates of ~8 points for primary hip replacement, ~6 points for revision hip replacement, ~7 points for primary knee replacement and ~6 points for revision knee replacement. The $MIC_{adjusted}$ method used here was based on a predictive-modelling approach, designed to correct for the high prevalence of improved patients after joint replacement [17]. There is currently no consensus as to whether this represents a gold-standard. We have shown here that the $MIC_{adjusted}$ method appears to be the least conservative of the anchor-based methods, with estimates based on ROC or predictive-modelling without correction up to two points higher.

The estimates produced in this study are an update to those produced previously by our group using an earlier version of the NHS PROMs dataset [15]. The current study has several advantages, including the calculation of new estimates for elective, revision joint replacement and a larger, more contemporary patient sample, resulting in greater precision. We are not the first group to use predictive-

modelling to investigate the MIC of the OKS for primary knee replacement. Ingelsrud et al [26] previously reported similar MIC estimates to those found here: 9 points (95% CI 6-15) using ROC-based methodology and 8 points (95% CI 6-9) using a predictive-modelling approach. We are not aware of previous studies investigating the OHS/OKS to have adjusted for the prevalence and severity of disease in this population in order to account for spectrum bias [17].

Whilst we have also presented distribution-based MIC estimates in this study, our views align with those of the COSMIN group that these estimates are less clinically meaningful than anchor-based estimates. Distribution-based estimates can be considered to reflect statistical properties of the sample rather than changes that are necessarily important to patients [14]. This point is worth highlighting when it comes to interpretation of the minimal detectable change. The MDC_{90} estimate is considered to be the change in score in either direction that is beyond the measurement error of the instrument. In this study, we found MDC_{90} estimates of ~6-7 points, which are higher than in our previous study [15]. Perhaps the main relevance of this finding concerns the interpretation of $MIC_{adjusted}$ estimates for revision hip and knee replacement, where the estimates were smaller than the MDC_{90} . To reconcile this, we support the interpretation of Terwee [27] that anchor-based estimates (which patients have identified to be important) should not be altered based on the measurement error of the instrument.

Our study does have some potential methodological limitations that should be considered. The use of any anchor-based estimate is reliant on the ability of the anchor question to identify a meaningful change in clinical state. We have interpreted a *minimal* important change as being ‘a little better’ following surgery. However, others have argued that patients may need to experience a greater transition for surgery to have been perceived as worthwhile [28]. The timing of the post-operative PROM administration in surgery is often challenging: it must be late enough to capture final status, yet early enough such that the patient still recalls their pre-operative status. In this study, we found that the anchor question was most highly correlated to the post-operative score, which is evidence for recall bias. This problem does not have an easy solution, since earlier administration of the post-operative PROM may capture patients still recovering from surgery. We believe that PROM administration at 6 months is reasonable and recent evidence suggests that it provides an accurate reflection of joint function at 5 years [29].

When interpreting the estimates from this study, it is important to appreciate that meaningful changes and differences are not invariable characteristics and may be affected by a number of different factors. These may include patient characteristics (such as age, gender, socio-economic status) [28,30]; disease characteristics (such as severity or prevalence [17]); procedure characteristics (for example, ‘minor’ versus ‘major’ revision procedures); and health-

care system factors (such as waiting times or insurance status). As such, the estimates produced in this study may benefit from cross-validation in other patient populations. The use of predicted-modelling based approaches to calculate the MIC provides the potential for covariate adjustment. This may allow better tailoring of MIC estimates to individual patients or groups that do not have average characteristics. It is important to note that predictive-modelling techniques remain vulnerable to problems with floor and ceiling effects within an instrument and new methods are being developed to account for these [31,32].

For any study where a PROM is chosen to measure the outcome, it is important to ensure that the instrument has been validated for the construct of interest. Both the OHS and OKS have been validated for primary and revision joint replacement [7,8,12,13]. Whilst there is good evidence that these instruments measure changes in pain and function during activities of daily living [9,10], not all patients will be undergoing surgery for improvement in these domains alone. This limitation is especially important to consider in revision joint replacement, which may have a different goal from surgery (such as infection eradication) and, in these circumstances, a more appropriate instrument may need to be identified. Indeed, in the future, we may move towards “consensus-based” approaches where individual patients define the domains and values for meaningful change themselves [33].

In conclusion, this study has calculated contemporary estimates of clinically important changes and differences for the Oxford Hip and Knee Scores for primary and revision joint replacement. For clinical trials comparing two or more groups, the MID was calculated to be ~5 points. This estimate can be used to inform future sample size calculations. To measure changes over time for a single group of patients, the mean change method should be used, with 12.4 points needed to interpret improvement after primary hip replacement, 8.6 points for revision hip replacement, 10.5 points for primary knee replacement and 9.4 points following revision knee replacement. For change at the individual patient level, the MIC_{adjusted} estimate was ~8 points for primary hip replacement, ~6 points for revision hip replacement, ~7 points for primary knee replacement and ~6 points for revision knee replacement. Changes over time need to be interpreted in the context of the measurement error of the instrument, which was calculated to be ~6 points for primary hip and knee replacement and revision knee replacement and ~7 points for revision hip replacement.

Acknowledgements

We would like to acknowledge all patients who contributed data to the NHS PROMs programme over the study period.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.jclinepi.2021.12.016](https://doi.org/10.1016/j.jclinepi.2021.12.016).

CRedit authorship contribution statement

Shiraz A. Sabah: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Software, Writing – original draft. **Abtin Alvand:** Investigation, Writing – review & editing. **David J. Beard:** Conceptualization, Funding acquisition, Investigation, Methodology, Supervision, Writing – review & editing. **Andrew J. Price:** Conceptualization, Funding acquisition, Investigation, Methodology, Supervision, Writing – review & editing.

References

- [1] NHS Digital. Patient Reported Outcome Measures (PROMs) n.d.
- [2] Beard DJ, Davies LJ, Cook JA, MacLennan G, Price A, Kent S, et al. The clinical and cost-effectiveness of total versus partial knee replacement in patients with medial compartment osteoarthritis (TOP-KAT): 5-year outcomes of a randomised controlled trial. *Lancet* (London, England) 2019;394:746–56. doi:[10.1016/S0140-6736\(19\)31281-4](https://doi.org/10.1016/S0140-6736(19)31281-4).
- [3] Murray DW, MacLennan GS, Breeman S, Dakin HA, Johnston L, Campbell MK, et al. A randomised controlled trial of the clinical effectiveness and cost-effectiveness of different knee prostheses: The Knee Arthroplasty Trial (KAT). *Health Technol Assess* 2014;18:1–235 vii–viii. doi:[10.3310/hta18190](https://doi.org/10.3310/hta18190).
- [4] Sabah SA, Alvand A. The last word on placebo-controlled surgical trials. *Journal of Trauma and Orthopaedics* 2021.
- [5] NHS Digital. Patient Reported Outcome Measures (PROMs) in England - A guide to PROMS methodology. 2017.
- [6] Zotov E, Hills AF, de Mello FL, Aram P, Sayers A, Blom AW, et al. JointCalc: A web-based personalised patient decision support tool for joint replacement. *Int J Med Inf* 2020;142:104217. doi:[10.1016/j.ijmedinf.2020.104217](https://doi.org/10.1016/j.ijmedinf.2020.104217).
- [7] Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perceptions of patients about total hip replacement. *J Bone Joint Surg Br* 1996;78:185–90. doi:[10.1302/0301-620x.78b2.0780185](https://doi.org/10.1302/0301-620x.78b2.0780185).
- [8] Dawson J, Fitzpatrick R, Murray D, Carr A. Questionnaire on the perceptions of patients about total knee replacement. *J Bone Joint Surg Br* 1998;80:63–9. doi:[10.1302/0301-620X.80B1.7859](https://doi.org/10.1302/0301-620X.80B1.7859).
- [9] Harris K, Beard D, Dawson J, Price A. An exploratory and confirmatory analysis of the oxford hip score: Generation of sub-scales assessing self-reported function and pain. *Trials* 2013;14:1. doi:[10.1186/1745-6215-14-s1-p74](https://doi.org/10.1186/1745-6215-14-s1-p74).
- [10] Harris K, Dawson J, Doll H, Field RE, Murray DW, Fitzpatrick R, et al. Can pain and function be distinguished in the Oxford Knee Score in a meaningful way? An exploratory and confirmatory factor analysis. *Qual Life Res* 2013;22:2561–8. doi:[10.1007/s11136-013-0393-x](https://doi.org/10.1007/s11136-013-0393-x).
- [11] Harris K, Dawson J, Gibbons E, Lim C, Beard D, Fitzpatrick R, et al. Systematic review of measurement properties of patient-reported outcome measures used in patients undergoing hip and knee arthroplasty. *Patient Related Outcome Measures* 2016;7:101–8. doi:[10.2147/prom.s97774](https://doi.org/10.2147/prom.s97774).
- [12] Dawson J, Fitzpatrick R, Frost S, Gundle R, McLardy-Smith P, Murray D. Evidence for the validity of a patient-based instrument for assessment of outcome after revision hip replacement. *J Bone*

- Joint Surg - Series B 2001;83:1125–9. doi:[10.1302/0301-620X.83B8.11643](https://doi.org/10.1302/0301-620X.83B8.11643).
- [13] Sabah SA, Alvand A, Beard DJ, Price AJ. Evidence for the validity of a patient-based instrument for assessment of outcome after revision knee replacement: Evaluation of the Oxford Knee Score using the UK National PROMS Dataset. *Bone Joint J* 2021;103-B(4):627–34. doi:[10.1302/0301-620X.103B4.BJJ-2020-1560.R1](https://doi.org/10.1302/0301-620X.103B4.BJJ-2020-1560.R1).
- [14] Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018;27:1147–57. doi:[10.1007/s11136-018-1798-3](https://doi.org/10.1007/s11136-018-1798-3).
- [15] Beard DJ, Harris K, Dawson J, Doll H, Murray DW, Carr AJ, et al. Meaningful changes for the Oxford hip and knee scores after joint replacement surgery. *J Clin Epidemiol* 2015;68:73–9. doi:[10.1016/j.jclinepi.2014.08.009](https://doi.org/10.1016/j.jclinepi.2014.08.009).
- [16] Terluin B, Eekhout I, Terwee CB, De Vet HCW. Minimal important change (MIC) based on a predictive modeling approach was more precise than MIC based on ROC analysis. *J Clin Epidemiol* 2015;68:1388–96. doi:[10.1016/j.jclinepi.2015.03.015](https://doi.org/10.1016/j.jclinepi.2015.03.015).
- [17] Terluin B, Eekhout I, Terwee CB. The anchor-based minimal important change, based on receiver operating characteristic analysis or predictive modeling, may need to be adjusted for the proportion of improved patients. *J Clin Epidemiol* 2017;83:90–100. doi:[10.1016/j.jclinepi.2016.12.015](https://doi.org/10.1016/j.jclinepi.2016.12.015).
- [18] Murray DW, Fitzpatrick R, Rogers K, Pandit H, Beard DJ, Carr AJ, et al. The use of the Oxford hip and knee scores. *J Bone Joint Surg Br* 2007;89-B:1010–14. doi:[10.1302/0301-620X.89B8.19424](https://doi.org/10.1302/0301-620X.89B8.19424).
- [19] Devji T, Carrasco-Labra A, Qasim A, Phillips M, Johnston BC, Johnston BC, et al. Evaluating the credibility of anchor based estimates of minimal important differences for patient reported outcomes: Instrument development and reliability study. *BMJ* 2020;369:0. doi:[10.1136/bmj.m1714](https://doi.org/10.1136/bmj.m1714).
- [20] Norman GR, Sloan JA, Wywich KW. Interpretation of changes in health-related quality of life. *Med Care* 2003;41:582–92. doi:[10.1097/01.mlr.0000062554.74615.4c](https://doi.org/10.1097/01.mlr.0000062554.74615.4c).
- [21] King MT. A point of minimal important difference (MID): A critique of terminology and methods. *Expert Rev Pharmacoecon Outcomes Res* 2011;11:171–84.
- [22] Cronbach LJ, Meehl PE. Construct validity in psychological tests. *Psychol Bull* 1955;52:281.
- [23] Dawson J, Doll H, Coffey J, Jenkinson C. Oxford on behalf of the. Responsiveness and minimally important change for the Manchester-Oxford foot questionnaire (MOXFQ) compared with AOFAS and SF-36 assessments following surgery for hallux valgus. *Osteoarthritis Cartilage* 2007;15:918–31. doi:[10.1016/j.joca.2007.02.003](https://doi.org/10.1016/j.joca.2007.02.003).
- [24] Hope RM. Rmisc: Ryan Miscellaneous library 2013.
- [25] Thiele C. Cutpointr: Improved estimation and validation of optimal cutpoints in R. 2021.
- [26] Ingelsrud LH, Roos EM, Terluin B, Gromov K, Husted H, Troelsen A. Minimal important change values for the Oxford Knee Score and the Forgotten Joint Score at 1 year after total knee replacement. *Acta Orthopaedica* 2018;89:541–7. doi:[10.1080/17453674.2018.1480739](https://doi.org/10.1080/17453674.2018.1480739).
- [27] Terwee CB. Estimating minimal clinically important differences and minimal detectable change. *J Hand Surg* 2019;44:e1.
- [28] De Vet HC, Ostelo RW, Terwee CB, Van Der Roer N, Knol DL, Beckerman H, et al. Minimally important change determined by a visual method integrating an anchor-based and a distribution-based approach. *Qual Life Res* 2007;16:131–42.
- [29] Dainty JR, Smith T, Clark EM, Whitehouse MR, Price A, Macgregor A. Trajectories of pain and function in the first 5 years after hip and knee joint replacement: An analysis of patient reported outcome data from the National Joint Registry of England, Wales, Northern Ireland and the Isle of Man. *Bone & Joint Journal* 2021.
- [30] De Vet HCW, Foumani M, Scholten MA, Jacobs WCH, Stiggelbout AM, Knol DL, et al. Minimally important change values of a measurement instrument depend more on baseline values than on the type of intervention. *J Clin Epidemiol* 2015;68:518–24. doi:[10.1016/j.jclinepi.2014.07.008](https://doi.org/10.1016/j.jclinepi.2014.07.008).
- [31] Ward MM, Guthrie LC, Alba M. Dependence of the minimal clinically important improvement on the baseline value is a consequence of floor and ceiling effects and not different expectations by patients. *J Clin Epidemiol* 2014;67:689–96. doi:[10.1016/j.jclinepi.2013.10.025](https://doi.org/10.1016/j.jclinepi.2013.10.025).
- [32] Bjorner JB, Terluin B, Trigg A, Hu J, Brady KJS, Griffiths P. Establishing Thresholds for Meaningful Within-individual Change Using Longitudinal Item Response Theory 2021. doi:[10.21203/rs.3.rs-371137/v1](https://doi.org/10.21203/rs.3.rs-371137/v1).
- [33] Dekker J. The minimal clinically important difference re-considered. *Osteoarthritis Cartilage* 2019;27:1403–4. doi:[10.1016/j.joca.2019.05.014](https://doi.org/10.1016/j.joca.2019.05.014).