**Selected Proceedings from the 10th International Congress of Arthroplasty Registries
Guest Editor: Ola Rolfson MD, PhD**

How do Patient-reported Outcome Scores in International Hip and Knee Arthroplasty Registries Compare?

Lina Holm Ingelsrud PT, MSc, PhD[1] , J. Mark Wilkinson MB ChB, PhD[2], Soren Overgaard MD, PhD[3,4],
Ola Rolfson MD, PhD[5], Brian Hallstrom MD[6], Ronald A. Navarro MD[7], Michael Terner MSc[8],
Sunita Karmakar-Hore PT, MSc[8], Greg Webster MSc[8], Luke Slawomirski PT, MSc[9],
Adrian Sayers PhD, MSc, MSc(Dist), PGDip(LSHTM), BSc(Hons)[10], Candan Kendir MD, MPH[9],
Katherine de Bienassis MPH[9], Niek Klazinga MD[9], Annette W. Dahl PT, PhD[11], Eric Bohm MD, MSc[12]

Abstract

Background Patient-reported outcome measures (PROMs) are the only systematic approach through which the patient's perspective can be considered by surgeons (in determining a procedure's efficacy or appropriateness) or healthcare systems (in the context of value-based healthcare). PROMs in registries enable international comparison of patient-centered outcomes after total joint arthroplasty, but the extent to which those scores may vary between different registry populations has not been clearly defined.

Questions/purposes (1) To what degree do mean change in general and joint-specific PROM scores vary across arthroplasty registries, and to what degree is the proportion of missing PROM scores in an individual registry associated with differences in the mean reported change scores? (2) Do PROM scores vary with patient BMI across registries? (3) Are comorbidity levels comparable across registries, and are they associated with differences in PROM scores?

Methods Thirteen national, regional, or institutional registries from nine countries reported aggregate PROM scores for patients who had completed PROMs preoperatively and 6 and/or 12 months postoperatively. The requested aggregate PROM scores were the EuroQol-5 Dimension Questionnaire (EQ-5D) index values, on which score 1 reflects "full health" and 0 reflects "as bad as death." Joint-specific PROMs were the Oxford Knee Score (OKS) and the Oxford Hip Score (OHS), with total scores ranging from 0 to 48 (worst-best), and the Hip Disability and Osteoarthritis Outcome Score-Physical Function shortform (HOOS-PS) and the Knee Injury and Osteoarthritis Outcome Score-Physical Function shortform (KOOS-PS) values, scored 0 to 100 (worst-best). Eligible patients underwent primary unilateral THA or TKA for osteoarthritis between 2016 and 2019. Registries were asked to exclude patients with subsequent revisions within their PROM collection period. Raw aggregated PROM scores and scores adjusted for age, gender, and baseline values were inspected descriptively. Across all registries and PROMs, the reported percentage of missing PROM data varied from 9% (119 of 1354) to 97% (5305 of 5445). We therefore graphically explored whether PROM scores were associated with the level of data completeness. For each PROM cohort, chi-square tests were performed for BMI

distributions across registries and 12 predefined PROM strata (men versus women; age 20 to 64 years, 65 to 74 years, and older than 75 years; and high or low preoperative PROM scores). Comorbidity distributions were evaluated descriptively by comparing proportions with American Society of Anesthesiologists (ASA) physical status classification of 3 or higher across registries for each PROM cohort.

Results The mean improvement in EQ-5D index values (10 registries) ranged from 0.16 to 0.33 for hip registries and 0.12 to 0.25 for knee registries. The mean improvement in the OHS (seven registries) ranged from 18 to 24, and for the HOOS-PS (three registries) it ranged from 29 to 35. The mean improvement in the OKS (six registries) ranged from 15 to 20, and for the KOOS-PS (four registries) it ranged from 19 to 23. For all PROMs, variation was smaller when adjusting the scores for differences in age, gender, and baseline values. After we compared the registries, there did not seem to be any association between the level of missing PROM data and the mean change in PROM scores. The proportions of patients with BMI 30 kg/m^2 or higher ranged from 16% to 43% (11 hip registries) and from 35% to 62% (10 knee registries). Distributions of patients across six BMI categories differed across hip and knee registries. Further, for all PROMs, distributions also differed across 12 predefined PROM strata. For the EQ-5D, patients in the younger age groups (20 to 64 years and 65 to 74 years) had higher proportions of BMI measurements greater than 30 kg/m^2 than older patients, and patients with the lowest baseline scores had higher proportions of BMI measurements more than 30 kg/m^2 compared with patients with higher baseline scores. These associations were similar for the OHS and OKS cohorts. The proportions of patients with ASA Class at least 3 ranged across registries from 6% to 35% (eight hip registries) and from 9% to 42% (nine knee registries).

Conclusion Improvements in PROM scores varied among international registries, which may be partially explained by differences in age, gender, and preoperative scores. Higher BMI tended to be associated with lower preoperative PROM scores across registries. Large variation in BMI and comorbidity distributions across registries suggest that future international studies should consider the effect of adjusting for these factors. Although we were not able to evaluate its effect specifically, missing PROM data is a recurring challenge for registries. Demonstrating generalizability of results and evaluating the degree of response bias is crucial in using registry-based PROMs data to evaluate differences in outcome. Comparability between registries in terms of specific PROMs collection, postoperative timepoints, and demographic factors to enable confounder adjustment is necessary to use comparison between registries to inform and improve arthroplasty care internationally.

Level of Evidence Level III, therapeutic study.

[1]Department of Orthopaedic Surgery, Copenhagen University Hospital Hvidovre, Hvidovre, Denmark

[2]Department of Oncology and Metabolism, The University of Sheffield, Sheffield, UK

[3]Department of Orthopaedic Surgery and Traumatology, Copenhagen University Hospital, Bispebjerg, Denmark

[4]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

[5]Department of Orthopaedics at Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

[6]Department of Orthopaedic Surgery, University of Michigan, Ann Arbor, MI, USA

[7]Kaiser Permanente, Harbor City, CA, USA

[8]Canadian Institute for Health Information, Toronto, ON, Canada

[9]Organisation for Economic Co-Operation and Development, Paris, France

[10]Musculoskeletal Research Unit, Learning & Research, University of Bristol, Southmead Hospital, Bristol, UK

[11]Orthopedics Department of Clinical Sciences Lund, Lund University, Lund, Sweden

[12]Department of Surgery, University of Manitoba, Winnipeg, MB, Canada

L. H. Ingelsrud ✉, Department of Orthopaedic Surgery, Copenhagen University Hospital Hvidovre, Kettegaard Alle 30, 2650 Hvidovre, Denmark, Email: lina.holm.ingelsrud@regionh.dk

Introduction

Patient-reported outcome measures (PROMs) are increasingly emphasized in patient-centered and value-based healthcare because direct input from patients is necessary to measure outcomes that are considered important to patients. Numerous general and joint-specific PROMs have been introduced in clinical practice worldwide at the national, regional, or local level. The Swedish Hip Arthroplasty Register initiated nationwide PROM collection in 2002, and many other national-level registries have subsequently followed [1, 31]. PROM implementation in registries provides health professionals, hospitals, and national or administrative entities with data to monitor, inform, and improve patient-centered arthroplasty care.

International comparison is readily performed for more traditional outcomes such as revision rates, enabling up-to-date evaluations of prosthesis performance [12, 20]. Including PROMs in registries further enables international comparison of patient-oriented outcomes such as pain and physical function [37]. Differences in PROM scores can demonstrate problems with access to care, and they are used as early indicators of the effectiveness of care [4]. Variations in PROM collection (such as the instrument used and length of follow-up) can make meaningful comparisons across registries challenging. However, recent work by the International Society of Arthroplasty Registries (ISAR) PROM group found sufficient similarities in administration, quality assurance methods, and collected variables to enable international comparison [3].

The Organization for Economic Co-operation and Development's (OECD) biennial Health at a Glance report compared a range of healthcare quality indicators across OECD member countries, including the quality and outcomes of hip and knee arthroplasty for osteoarthritis [26, 27]. The two most recent editions (2019 and 2021) reported change in scores for common generic and condition-specific PROMs across registries. Using observational data from registries to compare treatment outcome requires careful consideration because of the risks of model misspecification, confounding, bias, and chance [25, 33]. Although the OECD reports used direct standardization to adjust for possible confounding by age, gender, and preoperative PROM scores, a number of additional factors are important to consider in hip and knee arthroplasty research, such as patient demographics, socioeconomic status, joint-specific history, BMI, lifestyle factors, and comorbidities [30]. Although many registries collect such case-mix variables, differences in definition and categorization hinder comparison across registries [3]. Two important confounders that may be relevant and feasible to collect across registries are BMI and comorbidities. One additional but vital challenge for registry-based comparison is the potential response bias that arises from missing PROM data.

Ensuring high response rates for PROMs in registries is costly, and differences in survey logistics may explain the large variance in registries PROM completeness [3]. The question is whether research results, reflecting patients with complete PROMs, are generalizable to the larger population undergoing arthroplasty, which calls for an evaluation of the consequence of varying degree of PROM completeness across registries.

We therefore asked: (1) To what degree do mean change in general and joint-specific PROM scores vary across arthroplasty registries, and to what degree is the proportion of missing PROM scores in an individual registry associated with differences in the mean reported change scores? (2) Do PROM scores vary with patient BMI across registries? (3) Are comorbidity levels comparable across registries, and are they associated with differences in PROM scores?

Materials and Methods

Data Collection

Emailed invitations were sent to all OECD members involved in the Patient-Reported Indicator Survey (PaRIS) initiative and all registries affiliated with ISAR. These registries are a mix of national, regional, or single institutional coverage. For this study, 13 national, regional, or institutional registries in nine countries agreed to submit data (Table 1). One aim of the ISAR PROM working group is to promote harmonization of PROM collection to allow direct comparison within and between registries over time. The rationale for inviting broadly and analyzing data that originate from different registry types was to increase the number of registries with comparable datasets. Furthermore, this strategy enabled us to evaluate and elucidate some challenges that arise when performing international comparison of PROM outcome.

Registries reported details of their data collection, descriptive patient characteristics, and PROM scores in a standardized Excel spreadsheet. Descriptive patient characteristics were reported for patients comprising separate cohorts for general PROMs for the hip (Supplementary Table 1; http://links.lww.com/CORR/A848) and knee (Supplementary Table 2; http://links.lww.com/CORR/A849) registries and hip- (Supplementary Table 3; http://links.lww.com/CORR/A850) and knee-specific (Supplementary Table 4; http://links.lww.com/CORR/A851) PROMs, separately. We requested data for patients who underwent surgery from 2016 through 2019 and who had completed PROMs both preoperatively and at 6 months and/or 12 months postoperatively (Table 1). The three most recent years of surgery with fully completed PROM collection were considered reasonable to capture sufficient sample sizes and to present contemporary results. To calculate response rates, we

Table 1. Registry design and type of PROM administration

Country	Registry	Coverage	Years of surgery for reported PROM data	Patients an response rates of reported generic and specific PROMs[a]						Survey type	Postoperative PROM administration	BMI available	Type of comorbidity indicator
				EQ-5D Hip	EQ-5D Knee	HOOS-PS	KOOS-PS	OHS	OKS				
Australia	Australian Orthopaedic Association National Joint Replacement Registry	National	July 2018 to December 2019	70% (3150 of 4505)[b]	69% (4325 of 6268)[b]			69% (3097 of 4505)	68% (4249 of 6268)	Electronic and/or telephone	6 months	Yes	ASA
Canada	Alberta Bone and Joint Health Institute	Regional	2016 to 2018	11% (1141 of 10,187)[b,c]	9% (1349 of 14,730)[b,c]					Paper or electronic	12 months	No	None
Canada	Manitoba	Regional	2016 to 2018	41% (1514 of 3661)[d]	37% (2020 of 5411)[c]			46% (1677 of 3661)	41% (2241 of 5411)	Paper	12 months	Yes	None
England	National Health Service England National PROMs Programme	National	January 2016 to December 2018	37% (73,965 of 199,982)[e]	35% (79,678 of 226,514)[e]			40% (79,848 of 199,982)	38% (85,281 of 226,514)	Paper	6 months	Yes	ASA
Finland	Coxa	Regional	2017 to 2019					51% (2130 of 4144)	46% (2225 of 4828)	Paper or electronic	12 months	Yes	ASA
Ireland	Irish National Orthopaedic Registry	National	January 2016 to December 2018	88% (1389 of 1577)[b]	88% (1013 of 1156)[b]			91% (1599 of 1762)	91% (1235 of 1354)	Paper or electronic	6 months	Yes	ASA
Italy	IRCCS Galeazzi Orthopedic Instritute	Single center	2016 to 2018	47% (400 of 858)[d]	51% (299 of 592)[d]	45% (388 of 858)	49% (289 of 592)			Paper (preoperative) and electronic or phone (postoperative)	6 and 12 months	Yes	ASA
Italy	Rizzoli Orthopedic Institute	Single center	2019	56% (342 of 616)[e]	56% (138 of 248)[e]	55% (341 of 616)	55% (137 of 248)			Paper	6 and 12 months	Yes	ASA
Italy	Tuscany; Orthopaedic PROMs and PREMs Observatory	Regional	2018 to 2019					3% (140 of 5445)		Electronic	6 and 12 months	Yes	None
The Netherlands	Landelijke Registratie Orthopedische Implantaten	National	2016 to 2018	35% (25,643 of 73,085)[e]	31% (20,831 of 66,874)[e]	32% (23,400 of 73,085)	30% (20,194 of 66,874)	32% (23,148 of 73,085)	28% (18,623 of 66,874)	Electronic and/or paper	12 months	Yes	ASA

Table 1. continued

Country	Registry	Coverage	Years of surgery for reported PROM data	Patients an response rates of reported generic and specific PROMs[a]						Survey type	Postoperative PROM administration	BMI available	Type of comorbidity indicator
				EQ-5D Hip	EQ-5D Knee	HOOS-PS	KOOS-PS	OHS	OKS				
Sweden	The Swedish Knee Arthroplasty Register PROM program	Regional	2016 to 2018		73% (3239 of 4422)[e]		70% (3117 of 4422)			Paper	12 months	Yes	ASA and Charnley[f]
Sweden	The Swedish Hip Arthroplasty Register	National	2016 to 2018	70% (21,668 of 30,919)[g]						Paper	12 months	Yes	ASA and Charnley[f]
Switzerland	Geneva Arthroplasty Registry	Single center	2016 to 2018	60% (300 of 499)[d]	64% (241 of 374)[d]					Paper	12 months	Yes	ASA and Charnley

[a]Response rates are calculated as the number of patients for which data was aggregated (numerator) and the number of patients meeting inclusion and exclusion criteria for which PROM surveys were offered (denominator). If both 6- and 12-month data were available, we included 12-month data in the analyses.
[b]EQ-5D 5L version used.
[c]EQ-5D VAS scale not available.
[d]EQ-5D index crosswalked from SF-12 version 1 (Italy Galeazzi) or version 2 (Canada Manitoba and Switzerland Geneva).
[e]EQ-5D 3L version used.
[f]Data available for a subset of patients.
[g]EQ-5D 3L used until 2016; 5L used from 2017.

requested the number of patients for which data were aggregated (numerator), and additionally the number of patients meeting inclusion and exclusion criteria for which PROM surveys were offered (denominator). The reported proportions with complete PROM data at both the preoperative and postoperative timepoints varied across registries and PROMs from 3% (140 of 5445) to 91% (1235 of 1354) (Table 1).

Patients

Inclusion criteria were patients aged 20 years or older with a principal diagnosis of osteoarthritis undergoing elective primary and unilateral THA or TKA. Exclusion criteria were patients with subsequent (including both subsequent contralateral primary arthroplasty and revision of the initial surgery) hip arthroplasty for hip registries, subsequent knee arthroplasty for knee arthroplasty registries, or death during the follow-up period (between surgery and the postoperative survey).

Outcomes Tools and Variables

The aggregation form included general health (EuroQol-5 Dimension questionnaire [EQ-5D] and the 12-item Short Form Survey [SF-12]) and condition-specific PROMs (Oxford Hip Score [OHS], Oxford Knee Score [OKS], Hip Disability and Osteoarthritis Outcome Score–Physical Function shortform [HOOS-PS], and Knee Injury and Osteoarthritis Outcome Score–Physical Function shortform [KOOS-PS]). The three-level EQ-5D index values calculated with the US valuation set were used in the analyses [34]. An index value of 1 reflects "full health" and 0 reflects "as bad as death." Minimum important change (MIC) values, reflecting the cutoff for improvement that is considered important to patients, for the EQ-5D index were suggested to be 0.3 after hip arthroplasty [28] and 0.1 after knee arthroplasty [23]. Registries collecting the five-level version were requested to convert these into the three-level version using van Hout et al.'s [40] algorithm. Registries reporting the SF-12 were instructed to map these values to the EQ-5D using the methods published by Sullivan et al. [39] (version 1) or Le [19] (version 2). The 12-item OHS and OKS reflect pain and joint function and are scored on a scale from 0 to 48 (worst to best) [7, 8]. The five-item HOOS-PS and seven-item KOOS-PS are short measures of physical function that are scored from 0 to 100 (worst to best) [6, 29]. Reported MIC values for these PROMs were 7.6 for the OHS and 6.9 for the OKS after primary hip and knee replacement, respectively [32]. An MIC value of 23 for the HOOS-PS after hip arthroplasty was suggested [28], whereas for the KOOS-PS, we found no reported MIC

value for knee arthroplasty specifically, but a value of 2.2 was suggested for patients with knee osteoarthritis [36]. These PROMs are the most commonly collected by arthroplasty registries [3].

We requested aggregate (mean and standard error) PROM scores for all eligible patients, as well as aggregate PROM scores by strata based on age (three strata: 20 to 64 years, 65 to 74 years, and 75 + years), gender (two strata: men and women), and baseline scores (two strata based on median values: EQ-5D hip: 0.52, EQ-5D knee: 0.59, OHS: 18.0, OKS: 19.0, HOOS-PS: 46.1, and KOOS-PS: 51.2 using data provided by the Dutch Arthroplasty Register and published medians) [24].

We also requested descriptive BMI distributions as the number of patients within a BMI category: ≤ 18.5 kg/m^2, 18.5 to ≤ 25 kg/m^2, > 25 to ≤ 30 kg/m^2, > 30 to ≤ 35 kg/m^2, > 35 to ≤ 40 kg/m^2, and > 40 kg/m^2 for each PROM stratum. This information was available for 12 of the included registries.

Finally, we requested available distributions of comorbidity levels for each PROM cohort collected by the registry, specifically for distributions across American Society of Anesthesiologists (ASA) physical status classification categories 1 to 5; the Charlson comorbidity index, categorized into scores 0, 1 to 2, 3 to 4, and 5 or more; and, the Charnley classification A, B1, B2, or C. Nine hip and nine knee registries reported ASA class distributions, and three registries additionally reported Charnley classifications (Table 1). No registries reported the Charlson comorbidity index.

Primary and Secondary Study Outcomes

Our primary study goal was to explore variations in generic and joint-specific PROM scores from arthroplasty registries across the world. To achieve this, we pooled the aggregated scores provided by each registry and graphically evaluated aggregated change in PROM scores, as well as the preoperative and postoperative PROM scores. Additionally, we explored the variation in aggregate change in PROM scores by the reported response rate.

Our secondary study goals were to explore the variation in BMI distributions across registries, whether BMI levels were associated with the PROM scores, and to evaluate the variation in comorbidity distributions across the registry cohorts. We used a visual and descriptive approach to explore BMI and comorbidity levels.

Statistical Analysis

Reported aggregated PROM scores were tabulated, and the variation across registries was examined descriptively. We converted reported standard errors to SDs as $SD = SE \times \sqrt{N}$ [13]. A random-effects model was used
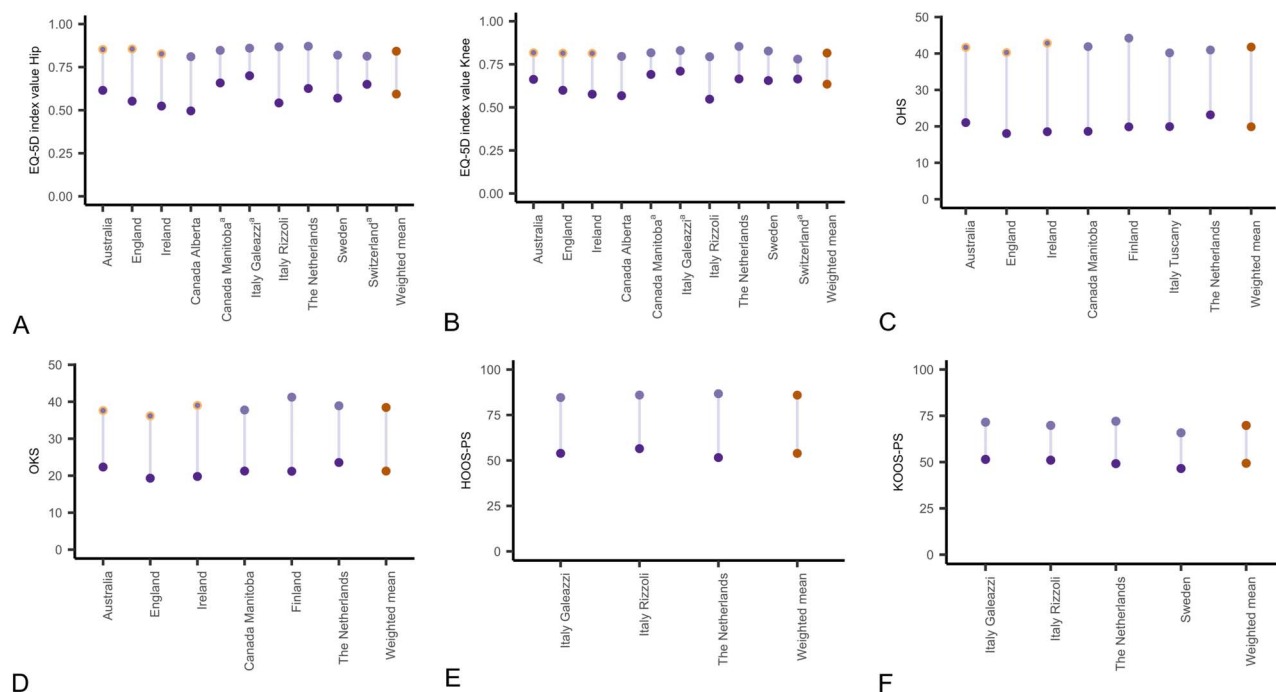
Fig. 1 A-F This figure shows the mean preoperative, 12-month postoperative, and change in (**A**) EQ-5D index values for hip arthroplasty registries, (**B**) EQ-5D index values for knee arthroplasty registries, (**C**) OHS for hip registries, (**D**) OKS for knee registries, (**E**) HOOS-PS for hip registries, and (**F**) KOOS-PS for knee registries. Orange circles denote registries with 6-month postoperative PROM administration. [a]Mapped from the SF-12.

to pool mean PROM scores because considerable heterogeneity between registries was expected. Statistical heterogeneity was calculated as the I^2 statistic. An I^2 value of 100% indicates maximal inconsistency between registry results [14]. Adjusted mean postoperative and change scores were generated using a direct standardization procedure to adjust for age, gender, and baseline score differences between registries. To explore whether registries with low PROM completeness could explain variation in PROM scores, we graphically investigated the response rate and change in PROM scores across registries. For the PROM cohorts where BMI was reported, we investigated the possible confounding factor of preoperative BMI on PROM outcomes by descriptively evaluating variation in distributions in separate BMI categories across registries. Comorbidity levels are presented descriptively to show which registries report which comorbidity index and their frequencies across PROM cohorts. Analyses were performed using either Microsoft Excel (Microsoft) or R version 4.1.0 (R Foundation).

Results

PROM Score Variation

Across all 13 registries, we found that unadjusted change scores were larger than published MIC values for all PROMs except the EQ-5D in hip registries, for which it

exceeded the MIC value of 0.3 in only 2 of 10 registries. For the EQ-5D, the pooled change in EQ-5D index values was 0.25 (95% confidence interval [CI] 0.20 to 0.29; I^2: 99.7%) for hip registries and 0.18 (95% CI 0.14 to 0.22; I^2: 99.5%) for knee registries. The mean change in the EQ-5D index ranged from 0.16 to 0.33 across hip registries and from 0.12 to 0.25 across knee registries. Pooled mean EQ-5D index values were 0.59 (95% CI 0.55 to 0.64) preoperatively and 0.84 (95% CI 0.83 to 0.86) postoperatively for hip registries and 0.63 (95% CI 0.59 to 0.67) preoperatively and 0.82 (95% CI 0.80 to 0.83) postoperatively for knee registries, with substantial heterogeneity for all models ($I^2 > 99\%$). This degree of statistical heterogeneity reflects the high degree of between-registry inconsistency in scores. Mean EQ-5D index values for individual registries (10 registries) ranged from 0.50 to 0.70 for hip registries (Supplementary Table 5; http://links.lww.com/CORR/A852) and from 0.55 to 0.71 for knee registries (Supplementary Table 6; http://links.lww.com/CORR/A853) preoperatively; they ranged from 0.81 to 0.87 (hip registries) (Fig. 1A) and from 0.78 to 0.85 (knee registries) (Fig. 1B) at 6 months to 12 months postoperatively. The change in the OHS and the OKS ranged from 18 to 24 and from 15 to 20, respectively. The pooled change in the OHS and the OKS was 22 (95% CI 20 to 24; I^2: 99.8%) and 17 (95% CI 15 to 19; I^2: 99.5%), respectively. The pooled mean OHS was 20 (95% CI 18 to
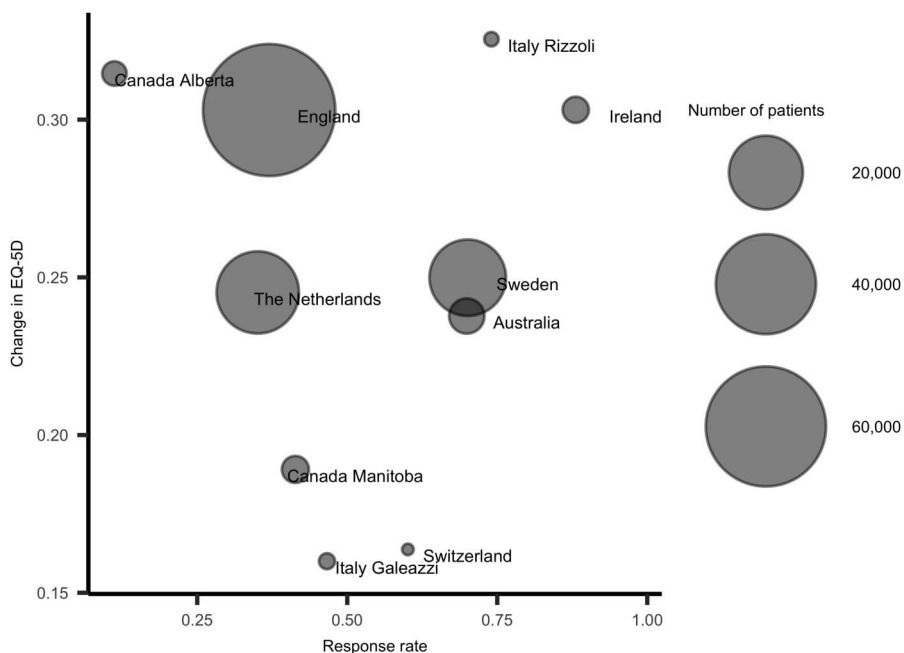
Fig. 2 This figure shows the association between change in EQ-5D index values and the response rates for the hip arthroplasty registries.

22; I^2: 99.9%) preoperatively and 42 (95% CI 40 to 43; I^2: 99.5%) postoperatively. The mean OHS for individual registries (seven registries) ranged from 18 to 23 preoperatively and from 40 to 44 postoperatively (Fig. 1C). The pooled mean OKS was 21 (95% CI 20 to 23; I^2: 99.9%) preoperatively and 38 (95% CI 37 to 40; I^2: 99.8%) postoperatively, and the mean OKS for individual registries (six registries) ranged from 19 to 24 preoperatively and from 36 to 41 postoperatively (Fig. 1D). The change in the HOOS-PS and KOOS-PS ranged from 29 to 35 and from 19 to 23, respectively. The pooled change in the HOOS-PS and KOOS-PS was 32 (95% CI 24 to 39; I^2: 95.8%) and 20 (95% CI 17 to 24; I^2: 97.9%), respectively. The pooled mean HOOS-PS was 54 (95% CI 48 to 60; I^2: 94.4%) preoperatively and 86 (95% CI 83 to 89; I^2: 75.8%) postoperatively. The mean HOOS-PS for individual registries (three registries) ranged from 52 to 57 preoperatively and from 85 to 87 postoperatively (Fig. 1E). The pooled mean KOOS-PS was 49 (95% CI 46 to 53; I^2: 97.2%) preoperatively and 70 (95% CI 65 to 74; I^2: 99.4%) postoperatively, and the mean KOOS-PS for individual registries (four registries) ranged from 47 to 51 preoperatively and from 66 to 72 postoperatively (Fig. 1F). The variation between registries was lower for all PROMs when scores were adjusted for age, gender, and baseline scores: EQ-5D for hip cohorts (Supplementary Table 5; http://links.lww.com/CORR/A852) and knee cohorts (Supplementary Table 6; http://links.lww.com/CORR/A853), OHS and HOOS-PS (Supplementary Table 7; http://links.lww.com/CORR/A854), and for OKS and

KOOS-PS (Supplementary Table 8; http://links.lww.com/CORR/A855). We found no association between the change in PROM scores and the response rates across registries, exemplified with the mean changes in the EQ-5D in hip cohorts (Fig. 2).

BMI Distributions

For the hip registries reporting EQ-5D and BMI data (nine registries), the proportion of patients with a BMI above 30 kg/m^2 ranged from 16% to 43%, with Australia, Canada, England, and Ireland having the largest proportions (Fig. 3A). Similar distributions were found for the registries that additionally reported BMI data for the OHS (seven registries) and HOOS-PS (three registries) cohorts (Supplementary Fig. 1; http://links.lww.com/CORR/A856). After we combined data from all registries and broke it down into the 12 predefined PROM strata, we found that patients in the younger age groups (20 to 64 years and 65 to 74 years) who had the lowest baseline scores had the highest BMI. These findings were evident for the EQ-5D, the OHS, and the HOOS-PS cohorts (Supplementary Fig. 2; http://links.lww.com/CORR/A857). For the knee registries, BMI was generally higher than for the hip registries. In the nine registries with EQ-5D data, proportions with a BMI above 30 kg/m^2 ranged from 35% to 62%, with Australia, Canada, Ireland, and England again exhibiting the highest proportions (Fig. 3B). Similar distributions were found for the OKS
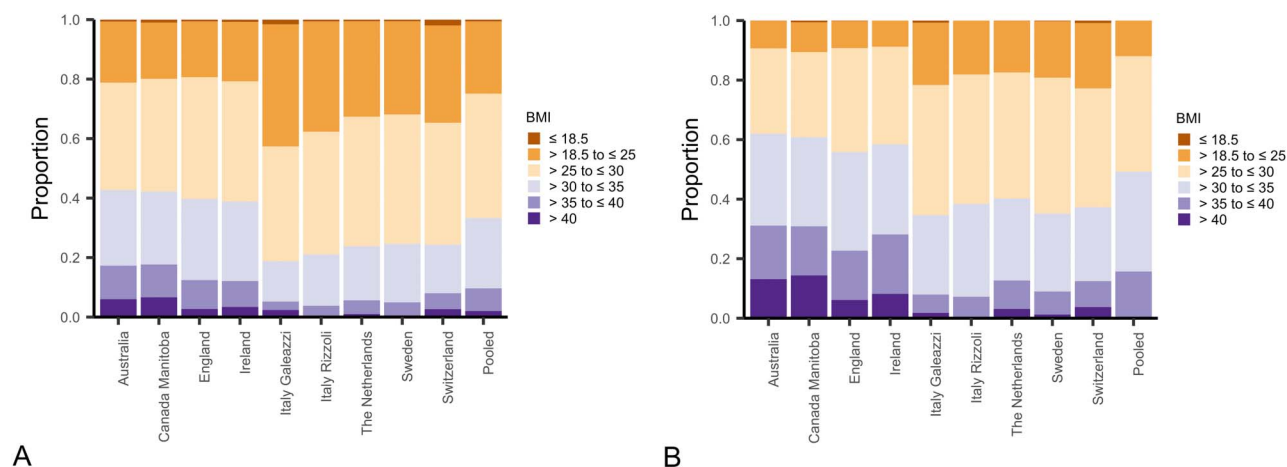
Fig. 3 This figure shows the proportions of patients within each BMI category across (**A**) hip and (**B**) knee registries with available BMI data for the EQ-5D cohorts.

(six registries) and KOOS-PS (four registries) cohorts (Supplementary Fig. 3; http://links.lww.com/CORR/A858). For all PROM cohorts, the youngest women with the lowest baseline scores had the highest BMIs (Supplementary Fig. 4; http://links.lww.com/CORR/A859).

Are Comorbidity Levels Comparable Across Registries?

Across all registries and PROM cohorts, proportions of patients with ASA Class 3 or above ranged from 6% to 35% for the hip registries and from 9% to 42% for the knee registries. Australia and Finland reported the largest proportions of patients with ASA Class 3 or higher for the hip (Supplementary Fig. 5; http://links.lww.com/CORR/A860) and knee cohorts (Supplementary Fig. 6; http://links.lww.com/CORR/A861).

Discussion

PROMs add an extra dimension to registries that have traditionally focused on implant survival and performance, and they are increasingly being collected in hip and knee arthroplasty registries. This study provides an initial effort to compare change in PROM scores from several registries worldwide. We gathered PROM data from registries across nine countries in Europe, North America, and Australia. We found that PROM scores across all registries reflected, on average, improved general health, decreased pain levels, and improved function for patients who received a hip or knee arthroplasty, supporting the widely acknowledged benefit of these procedures. However, variation in specific preoperative, postoperative, and change in PROM scores was observed. We did not find any apparent association between variation in PROM capture rate and mean change

in PROM scores across the registries studied. We also found differences in the distributions of BMI and comorbidity levels across the registries, and that BMI was associated with age, gender, and baseline PROM scores. These observations suggest that future international studies should consider the effect of adjusting for these factors when making cross-registry PROM comparisons.

Limitations

This study has several limitations. Firstly, local and national privacy regulations pose a major challenge for comparing PROM scores across multiple registries and nations because only aggregate-level data contribution is feasible. The aggregate-data format hinders regular confounder adjustment [38]. We performed a direct standardization procedure to account for age, gender, and preoperative scores, which required registries to break data into 12 strata. The strata needed to be fairly broad to avoid small cell counts. Adding potential confounder variables would have increased the number of PROM strata exponentially and would have simultaneously decreased the feasibility of data provision from registries because of privacy concerns about small cell counts. An alternative approach would be for each registry to perform confounder adjustment before submitting data, which presents other challenges including a larger workload for registries. Furthermore, whereas this study only focused on the potential effect of age, gender, preoperative PROM scores, BMI, and comorbidity, other factors such as work status, sociodemographic, and lifestyle factors are potentially important to consider, although they are challenging because of inconsistency in definitions and reporting across registries [30]. Because only raw PROM

scores and scores adjusted for age, gender, and baseline scores were presented in this study, a direct between-registry comparison of outcome is not feasible. However, the presented variation in scores is an important observation that warrants future investigation.

Another crucial question is whether it is valid to make direct comparisons of PROMs across healthcare systems and countries. The condition-specific PROMs used in this study (the OHS, OKS, KOOS-PS, and HOOS-PS) have been extensively evaluated in the hip and knee arthroplasty population [5, 11]. Contemporary evaluations of scale equivalence across language versions have not been performed, which complicates the comparativeness of the scores [22].

Further, although registries from three continents contributed data, other registries that were interested in contributing did not have resources to collate data or were hindered by data transfer rights, even though only aggregate-level data were required. The results therefore only reflect current PROM collection practices from certain parts of the world. In addition, we invited all types of registries, including small and local institutional registries, which are not comparable in scope, sampling strategy, or volume. The number of patients from whom the aggregated data were collected ranged widely across registries, from just above 100 patients to almost 80,000. However, the aim of our study was not to establish which countries have the most successful results, but rather to elaborate on the diversity between registries and possible challenges for transnational PROM outcome interpretation. Our pragmatic approach of inviting all registry types is appropriate for this descriptive study and allows for a discussion on factors that may impact on the generalizability and comparability of registry-based patient-relevant outcomes.

PROM Score Variation

Average change in PROM scores varied across registries. We found that all generic and condition-specific PROM scores across all registries improved from before to 6 months and 12 months after hip and knee arthroplasty. The improvement exceeded published thresholds for the MIC for all PROMs across all registries except the EQ-5D for hip registries, for which 8 of 10 registries had smaller average improvements than the MIC value of 0.3 [28, 32, 36]. Several studies have pointed at large variance in MIC values based on the derivation methodology [18], which could explain the relatively large discrepancy between the smaller MIC cutoff of 0.1 for knee arthroplasty and larger value of 0.3 for hip arthroplasty for the EQ-5D. Using MIC values as quality indicators in registries is not an established practice. Another strategy for evaluating the variation in improvements between registries could have been to compare proportions of patients in each registry

who improved more than published MIC values for the evaluated PROMs. However, such a comparison is complicated by the methodological challenges in selecting robust MIC cutoffs for the PROM under study. The variation in preoperative, postoperative, and change in PROM scores across registries may reflect ethnic and cultural differences between countries, such as surgical indication and access to surgery, as well as the real effectiveness of surgery or the healthcare system, and these differences are not easily quantified. Direct standardization of PROM scores, which reduced the between-registry variation, only partially considered patient differences [27].

The response rates varied largely between registries. Although we found no evidence that the degree of missing PROM data was associated with the average change in PROM scores, the aggregate data we had available did not enable an ideal analysis, and any possible response bias cannot be rejected. An example from the National Joint Registry of England, Wales, and Northern Ireland indicated a lower revision rate after shoulder arthroplasty for the cohort of patients having completed PROMs compared with those who had not (Fig. 4) [1]. That analysis clearly emphasizes that anyone using or publishing PROMs should make extensive efforts to consider response bias in the report and consider how missing data has arisen. The inability to rule out response bias in this study means the presented differences in PROM scores should be interpreted with caution, and captured PROM scores may reflect a more positive outcome of surgery than if all eligible patients undergoing surgery were included.

BMI Distributions

We observed differences in BMI distributions across registries. We found that higher BMI was associated with worse preoperative severity levels (particularly for younger patients) and for knee arthroplasty (particularly in women). Our findings are limited by the descriptive analyses performed based only on proportions of patients in separate BMI categories and PROM strata. Other studies have found conflicting evidence as to whether BMI influences the outcomes of hip and knee arthroplasty [10, 16, 17, 21]. However, a more recent study found that the change in the OHS was independent of BMI levels in patients undergoing hip arthroplasty when evaluated using a more rigorous statistical approach to consider bias inferred by floor and ceiling effects of the OHS model [33]. Our results show clear differences in preoperative PROM scores based on BMI, but whether BMI confounds the outcome of arthroplasty is still unanswered. Most registries were able to provide BMI data, either by direct collection in their registry or by linkage to
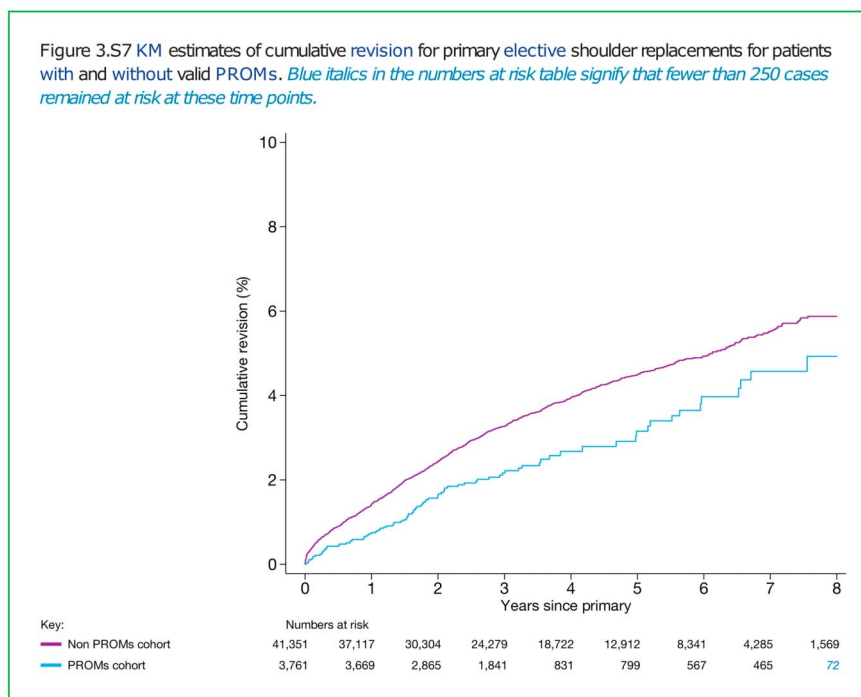
Fig. 4 This figure from the National Joint Registry of England, Wales, and Northern Ireland shows the cumulative revision of elective shoulder arthroplasty for patients with complete (blue) or missing (purple) PROMs. © National Joint Registry 2021 [1]. A color image accompanies the online version of this article.

other data sources, which enables future evaluations. Newer registries, or registries in the early implementation phase of PROM collection, can likewise improve their comparability with more established registries by ensuring incorporation of BMI data. Further analyses are needed to explore the confounding nature of BMI on the PROM scores, for which individual patient-level data is necessary.

Are Comorbidity Levels Comparable Across Registries?

The ASA class, reported by nine registries, was the only measure for which a comparison of comorbidities across registries was possible. We found that ASA class distributions varied greatly across reporting registries. The variation was similar to that in a recent study evaluating the association between ASA class and mortality in patients undergoing hip arthroplasty [35]. From the aggregate data we had, we were unable to evaluate the impact of ASA class on each registry's PROM scores. It has been suggested that comorbidity levels are important to consider when evaluating PROM scores after knee and hip arthroplasty [30]. Self-reported multimorbidity was shown to negatively impact the degree of improvement in the OHS and OKS after THA and TKA [41]. Further, when using the modified Charnley classification to evaluate comorbidities in patients undergoing knee arthroplasty,

Dunbar et al. [9] found that scores for measures of physical function, including the OKS, varied across Charnley classes. Furthermore, increasing comorbidity, as reflected by a worsening Charnley index scores, was associated with worsening physical function and pain after knee arthroplasty [15]. Measuring comorbidities is complex, and several comorbidity instruments have been developed for different purposes. The ASA class is a measure of disease severity levels as related to operative risk, and it can be used routinely in clinic [2]. From a registry perspective, longer and more comprehensive instruments, such as the Index of Coexistent Disease or the Functional Comorbidity Index, may be less feasible to implement. The ASA class was readily available for most registries in our study, and its influence on cross-registry PROM performance might be investigated further. Additionally, further studies should investigate whether the relatively simple and often readily available ASA class is a sufficient measure of comorbidity to further explore its potential confounding effect on the PROM scores.

Conclusion

Improvement in PROM scores varies internationally among hip and knee arthroplasty registries. The variation may be

partially explained by differences in age, gender, and pre-operative scores. Additionally, BMI and comorbidities may be relevant factors to adjust for when comparing registry-based PROM scores after THA and TKA. Future studies comparing PROMs across registries could investigate such confounders to determine whether adjustments are warranted and whether data are available from registries to feasibly make these adjustments. Our results must be seen in view of a risk of response bias that we were not able to fully evaluate with aggregate-level data.

Comparable data are crucial for using registry-based PROMs to inform differences in practices, facilitate learning, and improve arthroplasty care internationally. Newly established registries or registries initiating PROM collection can examine the data published here and previously published guidelines by the ISAR PROM working group [3] to ensure comparability of their data in terms of data collection timepoints and specific choice of PROMs. Importantly, differences in PROM scores may reflect differences in clinical practice, such as access to surgery and surgical indication as well as differences in treatment effects. When conducting observational analyses using registry-based PROM data to evaluate the effect of surgery, we urge researchers or health policy makers to evaluate the generalizability of the sample under study and consider response bias and how it may influence the results. One method for exploring response bias may be a clear description of the PROMs cohort by demographic characteristics at the time of surgery. This study illustrates some of the challenges involved in comparing PROMs from multiple registries and clearly establishes the need for comparable variables, both in terms of follow-up timepoints, choice of PROMs, cross-walking between PROMs, and preoperative patient demographics. Future international comparative analyses may ultimately serve to facilitate learning and improvement in arthroplasty care internationally.

References

1. Ben-Shlomo Y, Blom A, Boulton C, et al. The National Joint Registry 18th Annual Report 2021. Available at: https://www.ncbi.nlm.nih.gov/books/NBK576858/?report=classic. Accessed May 9, 2022.
2. Bjorgul K, Novicoff WM, Saleh KJ. Evaluating comorbidities in total hip and knee arthroplasty: available instruments. *J Orthop Traumatol*. 2010;11:203-209.
3. Bohm ER, Kirby S, Trepman E, et al. Collection and reporting of patient-reported outcome measures in arthroplasty registries: multinational survey and recommendations. *Clin Orthop Relat Res*. 2021;479:2151-2166.
4. Canadian Institute for Health Information, Organisation for Economic Co-operation and Development. OECD Patient-Reported Indicator Surveys (PaRIS) Initiative: patient-reported outcome measures (PROMs) for hip and knee replacement surgery international data collection guidelines. Available at: https://www.cihi.ca/sites/default/files/document/oecd-paris-hip-knee-data-collection-guidelines-en-web.pdf. Accessed January 15, 2022.
5. Collins NJ, Misra D, Felson DT, et al. Measures of knee function: International Knee Documentation Committee (IKDC) Subjective Knee Evaluation Form, Knee Injury and Osteoarthritis Outcome Score (KOOS), Knee Injury and Osteoarthritis Outcome Score Physical Function Short Form (KOOS-PS), Knee Outcome Survey Activities of Daily Living Scale (KOS-ADL), Lysholm Knee Scoring Scale, Oxford Knee Score (OKS), Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), Activity Rating Scale (ARS), and Tegner Activity Score (TAS). *Arthritis Care Res (Hoboken)*. 2011;63(suppl 11):S208-28.
6. Davis AM, Perruccio A V, Canizares M, et al. The development of a short measure of physical function for hip OA HOOS-Physical Function Shortform (HOOS-PS): an OARSI/OMERACT initiative. *Osteoarthritis Cartilage*. 2008;16:551-559.
7. Dawson J, Fitzpatrick R, Carr A, et al. Questionnaire on the perceptions of patients about total hip replacement. *J Bone Joint Surg Br*. 1996;78:185-190.
8. Dawson J, Fitzpatrick R, Murray D, et al. Questionnaire on the perceptions of patients about total knee replacement. *J Bone Joint Surg Br*. 1998;80:63-69.
9. Dunbar MJ, Robertsson O, Ryd L. What's all that noise? The effect of co-morbidity on health outcome questionnaire results after knee arthroplasty. *Acta Orthop Scand*. 2004;75:119-126.
10. Franklin PD, Li W, Ayers DC. The Chitranjan Ranawat award: Functional outcome after total knee replacement varies with patient attributes. *Clin Orthop Relat Res*. 2008;466:2597-2604.
11. Harris K, Dawson J, Gibbons E, et al. Systematic review of measurement properties of patient-reported outcome measures used in patients undergoing hip and knee arthroplasty. *Patient Relat Outcome Meas*. 2016;7:101-108.
12. Heckmann N, Ihn H, Stefl M, et al. Early results from the American Joint Replacement Registry: a comparison with other national registries. *J Arthroplasty*. 2019;34:S125-S134.e1.
13. Higgins J, Thomas J, Chandler J, et al. *Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021)*. Cochrane; 2021. Available at: https://training.cochrane.org/handbook/current/chapter-06#_Ref190897628. Accessed February 1, 2022.
14. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med*. 2002;21:1539-1558.
15. Hilton ME, Gioe T, Noorbaloochi S, et al. Increasing comorbidity is associated with worsening physical function and pain after primary total knee arthroplasty. *BMC Musculoskelet Disord*. 2016;17:1-10.
16. Jameson SS, Mason JM, Baker PN, et al. The impact of body mass index on patient reported outcome measures (PROMs) and complications following primary hip arthroplasty. *J Arthroplasty*. 2014;29:1889-1898.
17. Judge A, Batra RN, Thomas GE, et al. Body mass index is not a clinically meaningful predictor of patient reported outcomes of primary hip replacement surgery: prospective cohort study. *Osteoarthritis Cartilage*. 2014;22:431-439.
18. King MT. A point of minimal important difference (MID): a critique of terminology and methods. *Expert Rev Pharmacoecon Outcomes Res*. 2011;11:171-184.

19. Le QA. Probabilistic mapping of the health status measure SF-12 onto the health utility measure EQ-5D using the US-population-based scoring models. *Qual Life Res*. 2014;23:459-466.

20. Lewis PL, Tudor F, Lorimer M, et al. Short-term revision risk of patellofemoral arthroplasty is high: an analysis from eight large arthroplasty registries. *Clin Orthop Relat Res*. 2020;478:1222-1231.

21. Lübbeke A, Stern R, Garavaglia G, et al. Differences in outcomes of obese women and men undergoing primary total hip arthroplasty. *Arthritis Care Res*. 2007;57:327-334.

22. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res*. 2010;19:539-549.

23. Most J, Hoelen TA, Spekenbrink-spooren A, et al. Defining clinically meaningful thresholds for patient-reported outcomes in knee arthroplasty. *J Arthroplasty*. 2022;37:837-844.

24. National Health Service NHS England. National patient reported outcome measures (PROMs) programme consultation report. Available at: https://www.england.nhs.uk/wp-content/uploads/2017/10/proms-consultation-report.pdf. Accessed January 15, 2022.

25. Nørgaard M, Ehrenstein V, Vandenbroucke JP. Confounding in observational studies based on large health care databases: problems and potential solutions – a primer for the clinician. *Clin Epidemiol*. 2017;9:185-193.

26. Organisation for Economic Cooperation and Development. *Health at a Glance 2019: OECD Indicators*. OECD Publishing; 2019. Available at: https://doi.org/10.1787/4dd50c09-en. Accessed February 1, 2022.

27. Organisation for Economic Cooperation and Development. Health at a Glance 2021. *OECD Indicators*. OECD Publishing; 2021. Available at: https://doi.org/10.1787/19991312. Accessed February 1, 2022.

28. Paulsen A, Roos EM, Pedersen AB, et al. Minimal clinically important improvement (MCII) and patient-acceptable symptom state (PASS) in total hip arthroplasty (THA) patients 1 year postoperatively. *Acta Orthop*. 2014;85:39-48.

29. Perruccio AV, Lohmander L, Canizares M, et al. The development of a short measure of physical function for knee OA KOOS-Physical Function Shortform (KOOS-PS) - an OARSI/OMERACT initiative. *Osteoarthritis Cartilage*. 2008;16:542-550.

30. Rolfson O, Bohm E, Franklin P, et al. Patient-reported outcome measures in arthroplasty registries: report of the Patient-Reported Outcome Measures Working Group of the International Society of Arthroplasty Registries Part II. Recommendations for selection, administration, and analysis. *Acta Orthop*. 2016;87:9-23.

31. Rolfson O, Kärrholm J, Dahlberg LE, et al. Patient-reported outcomes in the Swedish Hip Arthroplasty Register: results of a nationwide prospective observational study. *J Bone Joint Surg Br*. 2011;93:867-875.

32. Sabah SA, Alvand A, Beard DJ, et al. Minimal important changes and differences were estimated for Oxford hip and knee scores following primary and revision arthroplasty. *J Clin Epidemiol*. 2021;143:159-168.

33. Sayers A, Whitehouse MR, Judge A, et al. Analysis of change in patient-reported outcome measures with floor and ceiling effects using the multilevel Tobit model: a simulation study and an example from a National Joint Register using body mass index and the Oxford Hip Score. *BMJ Open*. 2020;10:e033646.

34. Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. *Med Care*. 2005;43:203-220.

35. Silman AJ, Combescure C, Ferguson RJ, et al. International variation in distribution of ASA class in patients undergoing total hip arthroplasty and its influence on mortality: data from an international consortium of arthroplasty registries. *Acta Orthop*. 2021;92:304-310.

36. Singh JA, Luo R, Landon GC, et al. Reliability and clinically important improvement thresholds for osteoarthritis pain and function scales: a multicenter study. *J Rheumatol*. 2014;41:509-515.

37. Slawomirski L, van den Berg M, Karmakar-Hore S. Patient-Reported Indicator Survey (PaRIS): aligning practice and policy for better health outcomes. *World Medical Journal*. 2018;64:8-14. Available at: https://www.wma.net/wp-content/uploads/2018/10/WMJ_3_2018-1.pdf. Accessed February 1, 2022.

38. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*. 2016;355:1-7.

39. Sullivan PW, Ghuschshyan V. Mapping the EQ-5D Index from the SF-12: US general population preferences in a nationally representative sample. *Med Decis Making*. 2006;26:401-409.

40. van Hout B, Janssen MF, Feng YS, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value Health*. 2012;15:708-715.

41. Zhang L, Lix LM, Ayilara O, et al. The effect of multimorbidity on changes in health-related quality of life following hip and knee arthroplasty. *Bone Joint J*. 2018;100:1168-1174.